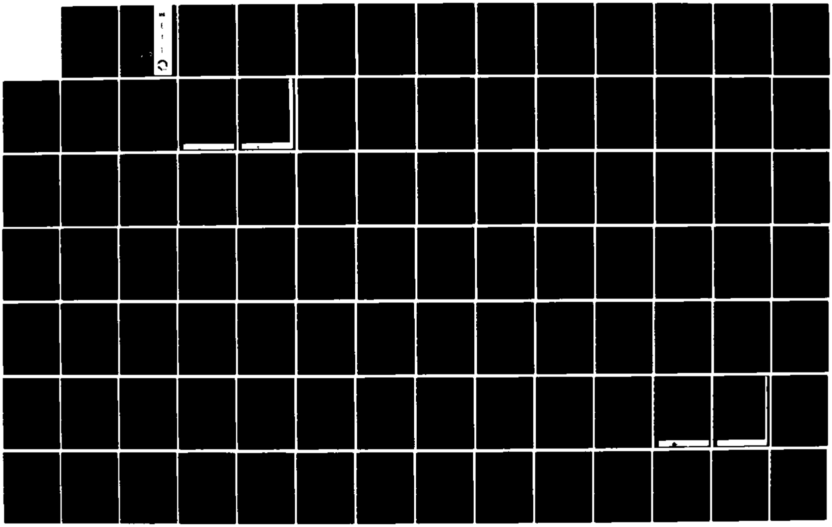
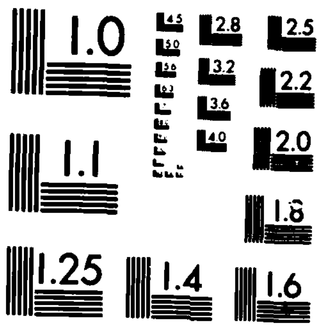
AD-A150 189 FEATURE EXTRACTION ASSESSMENT STUDY(U) ANALYTIC SCIENCES CORP READING MA M J CARLOTTO ET AL. NOV 84 ETL-0377 DACA76-82-C-0004 1/4

UNCLASSIFIED F/G 8/2 NL




1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

ETL-0377

AD-A150 189

Feature extraction assessment study, final report

Mark Carlotto
Victor Tom
Gregory Wallace

The Analytic Sciences Corporation
One Jacob Way
Reading, Massachusetts 01867

October 1984

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.

Prepared for

U.S. ARMY CORPS OF ENGINEERS
ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060-5546

E T L

U.S. ARMY ENGINEER TOPOGRAPHIC LABORATORIES
ETL

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ETL-0377	2. GOVT ACCESSION NO.: AD-A150189
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Feature Extraction Assessment Study, Final Report	5. TYPE OF REPORT & PERIOD COVERED: Final Report April 1982-August 1984
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): M. J. Carlotto, V. T. Tom, G. K. Wallace	8. CONTRACT OR GRANT NUMBER(s): DACA76-82-C-0004
9. PERFORMING ORGANIZATION NAME AND ADDRESS: The Analytic Sciences Corporation, One Jacob Way, Reading, MA 01867	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U.S. Army Engineer Topographic Laboratory, Leaf Road, Bldg. 2592, Fort Belvoir, VA 22060	12. REPORT DATE: November 1984
	13. NUMBER OF PAGES: 300
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Defense Mapping Agency, Bldg. 56; U.S. Naval Observatory, Washington, D.C. 20305	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Cartographic Feature Extraction,
Image Processing,
Pattern Recognition
Image Analysis,
Remote Sensing,

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of the feature extraction assessment study (FEAS) is to assess the degree to which the DMA feature extraction process can be automated by 1985. The major conclusion of the study is that given the state-of-the-art in image understanding and pattern recognition, a fully automated system for extracting DMA features cannot be built by 1985 using either conventional black and white or multi-spectral/multi-source imagery. Based on —

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. an assessment of feature extraction processes operating solely on black and white imagery, a concept of operation for a semi-automated system is formulated, however, which involves the use of machine vision techniques to support the image analyst in a synergistic fashion. An assessment of multi-spectral and multi-source (MS/MS) imagery is also performed to determine if its use would benefit the feature extraction process. A concept of operation for a system to extract DMA features in MS/MS imagery is then developed. It is shown that a key step in the process, the determination of surface material type, can be automated to a large degree given current technology. Moreover, it appears possible to employ 2-d image understanding techniques to infer the presence of certain kinds of DMA features once a surface material map has been derived. Preliminary results indicate the approach to be promising and in need of further examination.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

1. INTRODUCTION

1.1 BACKGROUND

The purpose of the Feature Extraction Assessment Study (FEAS) was to assess the degree of automation feasible for an FY85 DMA digital feature extraction system. The FEAS was partitioned into six tasks. Each is briefly reviewed below.

- Task 1 - Review of Available Input Data: The objective of this task was to identify and characterize all data which may be used to support semi-automated feature extraction in the FY85 timeframe. The approach was to consider three major classes of data: sensor-derived data, maps and charts, and derived digital data (e.g., DTED and DFAD). The output from this task was a determination of the ability of various forms of input data to satisfactorily resolve or represent DMA features.

- Task 2 - Review of DMA Data Base Features: The objective of this task was to select a set of DMA features amenable to semi-automated feature extraction. The approach was to characterize DMA features in terms of extractability, attributes (e.g., size, contrast, homogeneity, shadowing/obscuration characteristics). The output from this task was a characterization of DMA features in terms of extractability attributes for use in identifying which features are most amenable to semi-automated extraction.

- Task 3 - Review and Assessment of Black-and-White Feature Extraction Techniques: The objective of this task was to review and assess image processing, pattern recognition, and knowledge-based techniques

available in FY83 applicable to semi-automated feature extraction. The approach to this task was to assess the above techniques within the context of the Image Understanding (IU) paradigm (a model that suggests how an image must be interpreted to permit the inference of meaningful object properties). The output from Task 3 was an assessment of black-and-white feature extraction techniques and an identification of the applicability of these techniques to various classes of DLMS Level V features.

- Task 4 - Development of a Concept of Operations: The objective of this task was to formulate a conceptual operational environment for an interactive semi-automated feature extraction system. TASC's approach to Task 4 was to review current baseline operations in order to construct a functional sequence associated with the use of techniques identified under Task 3. The output was a definition of sequences of operations performed by both man and machine to accomplish semi-automated feature extraction.

- Task 5 - Review and Assessment of Multi-Spectral/Multi-Source (MS/MS) Feature Extraction Techniques: The objective of this task was to identify MS/MS image processing, segmentation, classification, and object identification techniques that are candidates for inclusion in a semi-automated feature extraction system, and to assess their utility in increasing the level of automation and performance of such a system. The approach was to apply the IU paradigm developed in Task 3 to the problem of extracting DMA features from MS/MS imagery. The output consisted of a review and assessment of the above techniques within the context of a FY85 feature extraction system.

- Task 6 - Development of a Concept of Operation for a Multi-Spectral/Multi-Source Feature Extraction System: The objective of this task was to develop a concept of operation based on the use of MS/MS imagery. The approach was to use the concept of operation developed in Task 4 as a baseline for semi-automated feature extraction, identifying processes which could be further automated by using MS/MS imagery. The output was a concept of operation describing the ways in which MS/MS imagery could be used within a semi-automated feature extraction system.

This document, the final report for the FEAS, provides an overview of the feature extraction process at DMA, a review and assessment of image processing, pattern recognition, and artificial intelligence techniques which may be of use in increasing the level of automation feasible in a FY85 feature extraction system, and a description of how the above techniques may be integrated within the operational environment at DMA for the purpose of extracting features of interest to DMA.

1.2 OVERVIEW OF FEATURE EXTRACTION PROCESS

Feature extraction at DMA consists of the extraction from source material, of non-terrain-elevation-related information. A complete Feature Extraction source package contains a control base in the form of orthophotos, control manuscripts, or maps or charts; aids to feature identification such as imagery (rectified and unrectified), maps, charts, and textual material; and other map and chart data such as names, boundary and contour manuscripts, electronic data and bathymetric data.

An overview of the feature extraction process is given in Fig. 1.2-1. This process has two goals:

- Generation of Digital Feature Analysis Data (DFAD)
- Production of Planimetric Compilation Manuscripts.

DFAD is a digital description of simulation-significant features, which is integrated with Digital Terrain Elevation Data (DTED) to produce Digital Landmass Simulation (DLMS) data. Planimetric compilation manuscripts contain data of interest to a particular map or chart product (e.g., buoys on a nautical chart, or roads, railroads, and bridges on a topographic map).

The feature extraction process is shown in greater detail in Fig. 1.2-2. DFAD production is shown by the series of functions on the midpart of the figure from "retrieval/assess/process source material" to "final review".

The first step in DLMS Planimetric Feature production is the identification and extraction of simulation-significant

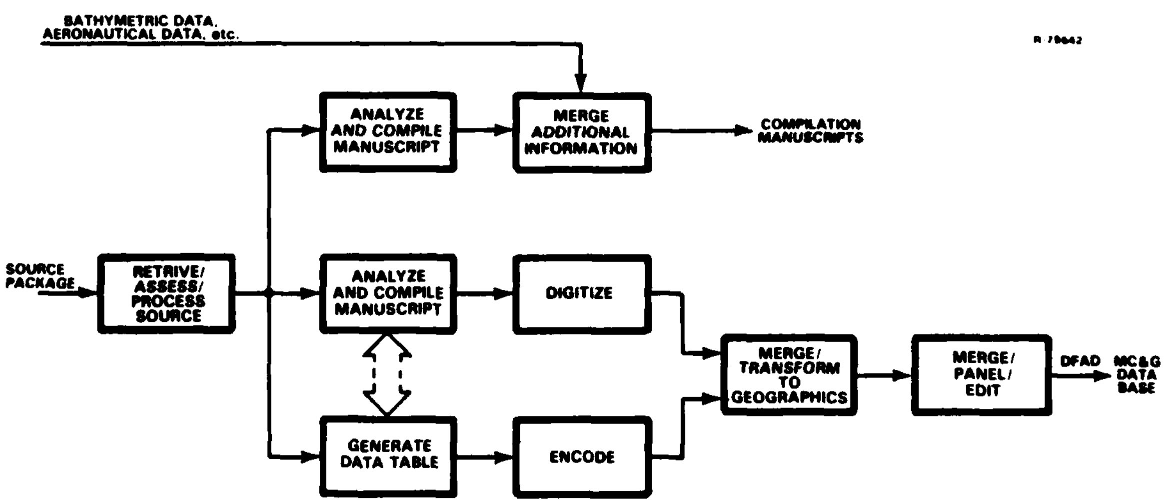
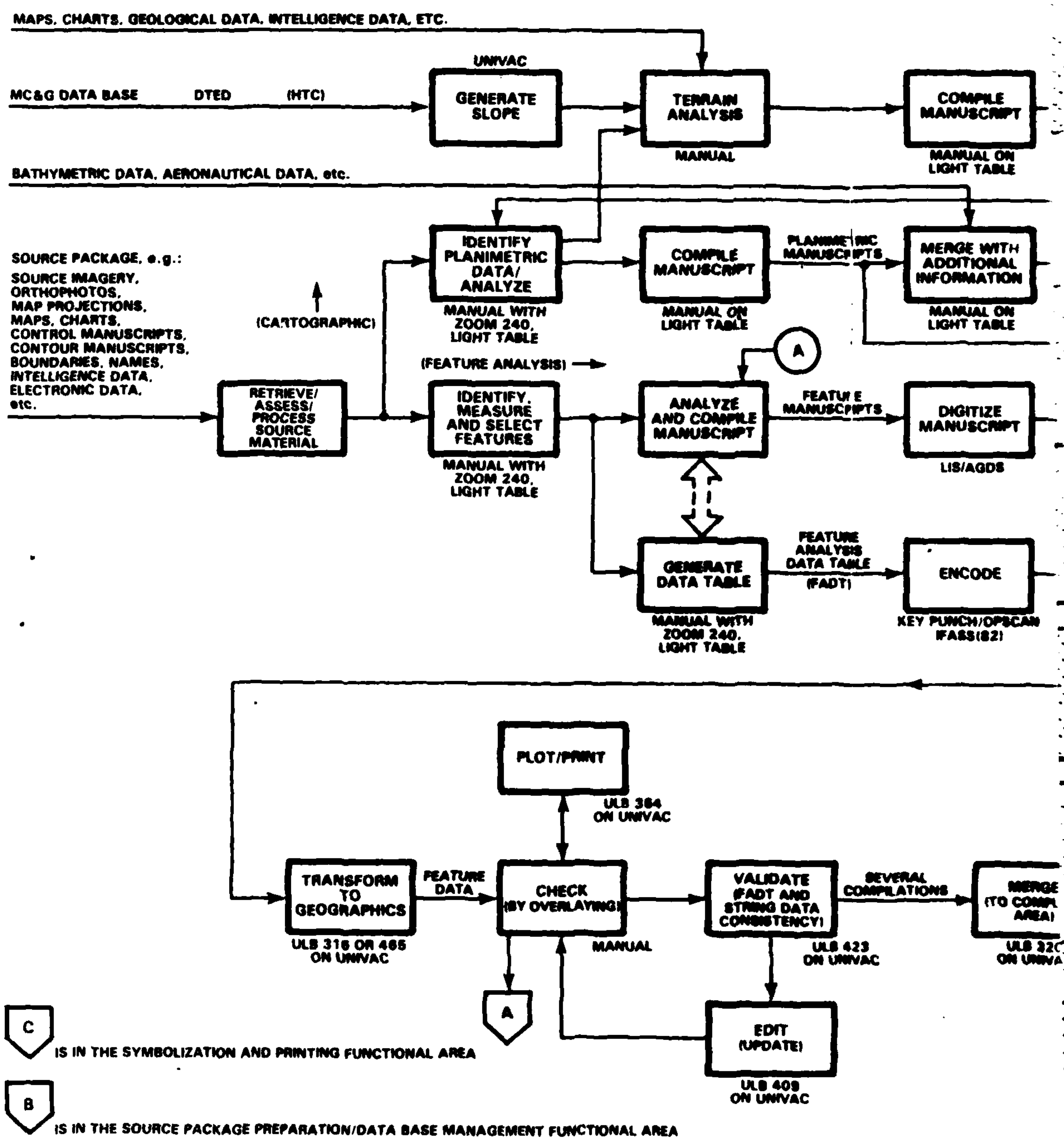
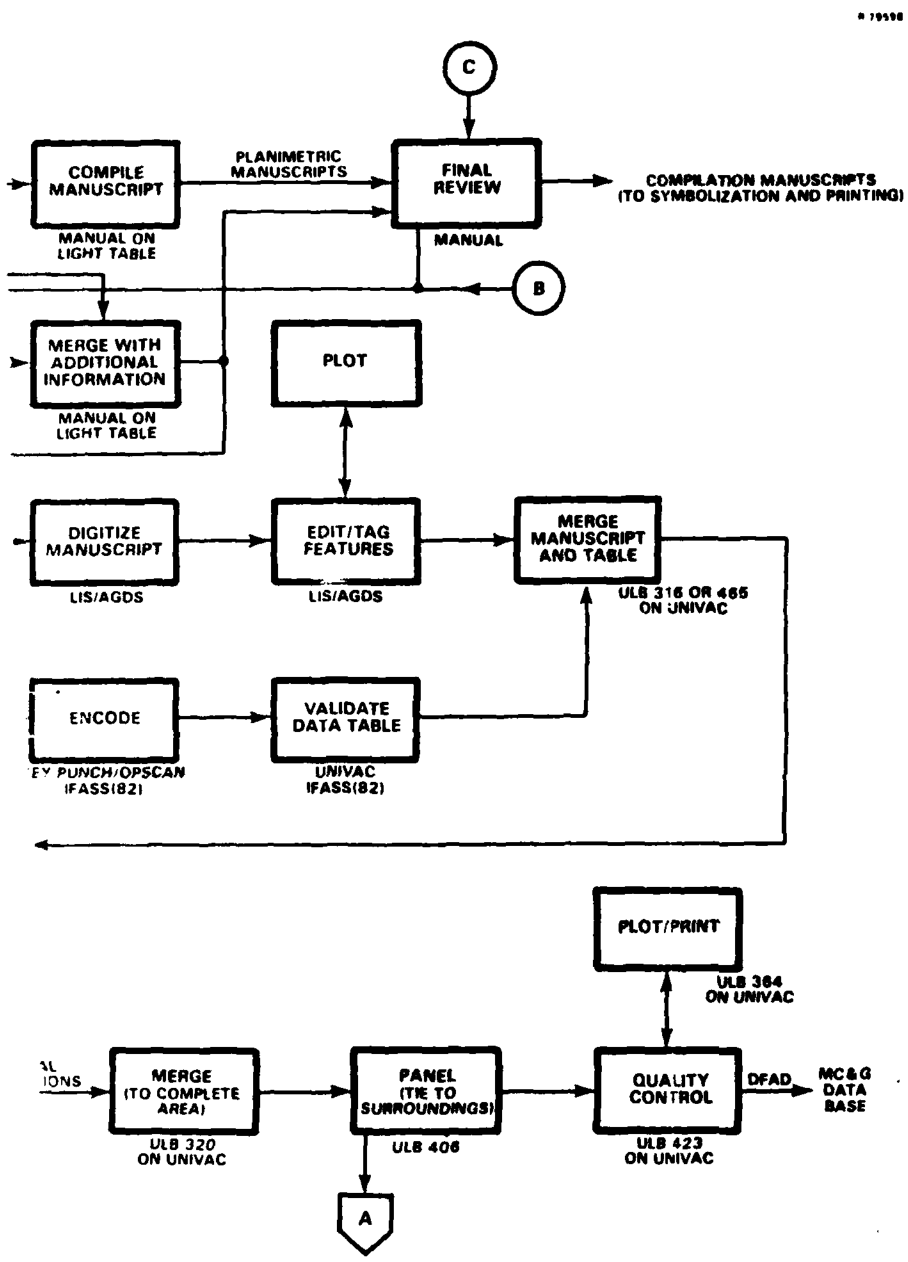


Figure 1.2-1 Feature Extraction Overview

TH


MAPS, CHARTS, GEOLOGICAL DATA, INTELLIGENCE DATA, ETC.
MC&G DATA BASE
DTED
(HTC)
UNIVAC
GENERATE SLOPE
TERRAIN ANALYSIS
MANUAL
COMPILE MANUSCRIPT
MANUAL ON LIGHT TABLE
BATHYMETRIC DATA, AERONAUTICAL DATA, etc.
SOURCE PACKAGE, e.g.:
SOURCE IMAGERY,
ORTHOPHOTOS,
MAP PROJECTIONS,
MAPS, CHARTS,
CONTROL MANUSCRIPTS,
CONTOUR MANUSCRIPTS,
BOUNDARIES, NAMES,
INTELLIGENCE DATA,
ELECTRONIC DATA,
etc.
(CARTOGRAPHIC)
IDENTIFY PLANIMETRIC DATA/ ANALYZE
MANUAL WITH ZOOM 240, LIGHT TABLE
COMPILE MANUSCRIPT
MANUAL ON LIGHT TABLE
PLANIMETRIC MANUSCRIPTS
MERGE WITH ADDITIONAL INFORMATION
MANUAL ON LIGHT TABLE
(FEATURE ANALYSIS)
A
RETRIEVE/ ASSESS/ PROCESS SOURCE MATERIAL
IDENTIFY, MEASURE AND SELECT FEATURES
MANUAL WITH ZOOM 240, LIGHT TABLE
ANALYZE AND COMPILE MANUSCRIPT
FEATURE MANUSCRIPTS
DIGITIZE MANUSCRIPT
LIS/AGDS
GENERATE DATA TABLE
MANUAL WITH ZOOM 240, LIGHT TABLE
FEATURE ANALYSIS DATA TABLE (FADT)
ENCODE
KEY PUNCH/OPSCAN IFASS(B2)
PLOT/PRINT
ULB 384 ON UNIVAC
TRANSFORM TO GEOGRAPHICS
ULB 316 OR 465 ON UNIVAC
FEATURE DATA
CHECK (BY OVERLAYING)
MANUAL
VALIDATE (FADT AND STRING DATA CONSISTENCY)
ULB 423 ON UNIVAC
SEVERAL COMPILATIONS
MERGE (TO COMP AREA)
ULB 32C ON UNIVA
A
EDIT (UPDATE)
ULB 409 ON UNIVAC
C IS IN THE SYMBOLIZATION AND PRINTING FUNCTIONAL AREA
B IS IN THE SOURCE PACKAGE PREPARATION/DATA BASE MANAGEMENT FUNCTIONAL AREA



2

Figure 1.2-2 Feature Extraction

features. Feature extraction is typically accomplished by photointerpreters (PIs) examining unrectified stereo imagery, feature by feature, using highly constrained search techniques. The identification and extraction of features is performed on a Zoom 240 stereoscope. Stereoscopic identification may be supported by other imagery, maps, and intelligence data. The identified features are manually delineated (i.e., drawn by hand) on a feature manuscript which may be a mylar overlay on an orthophoto or other control base.

Once delineated, each feature is also labeled according to a standardized feature identification (FID) code. The FID code is one of several informative types of entries called "Feature Descriptors" contained in the Feature Header Record. Another descriptor is the Feature Analysis Code (FAC) number, which provides a means for uniquely numbering each extracted feature in a manuscript region. The FAC number also refers to the feature boundary coordinates. A third Descriptor is called "Feature Type," which specifies the feature as lineal, areal, or point.

Of special importance is the fact that there are additional Descriptors which contain information about the physical properties of features beyond that of simple boundary information. This information is important for simulation purposes and makes the DLMS feature extraction problem more difficult than simple cartographic feature extraction. All of the information in the Feature Header Record, plus the coordinate boundary information, is called Digital Feature Analysis Data (DFAD). The additional descriptors include:

- Surface Material Category (SMC)
- Height

- Number of Structures
- Percent of Tree Coverage
- Orientation/Directivity
- Dimensions
- Roof Descriptor
- Shape Code
- Microdescriptor.

The general goal of the feature extraction process is to obtain, for each visually-identifiable feature within a given region of imagery:

1. the feature classification or name
2. the ordered (x,y,z) coordinates which define its boundary.

For DLMS planimetric features, the goal is to obtain the above items (the feature identification (FID) code and boundary coordinates), and to obtain all relevant Descriptors for each feature in the imagery.

Completed feature manuscripts are digitized on either the Advanced Graphic Display System (AGDS) or the Lineal Input System (LIS). The AGDS raster scans the manuscript and then performs a raster-to-vector conversion to obtain the data in the desired vector form. Using the LIS, an operator traces the manuscript features and the LIS captures the features in digital vector form. The digitized features are tagged with an identifying number common to the manuscript and the FADT. Both the AGDS and the LIS permit digitized data to be displayed and edited.

The FID code and associated feature descriptor data are also digitized. This is typically performed manually by a key-punch operator, but a scanning device (OPSCAN) may also be used.

Several improvements to the basic feature extraction process and system have recently been implemented or are planned. The Digital Interactive Multi-Image Analysis System (DIMIAS) is intended to aid DFAD production by classifying LANDSAT images, using multi-spectral pattern recognition, into surface material type categories. The classifications produced will be used to generate plots serving as baseline feature manuscripts for further manual processing.

The Interactive Feature Analysis Support System (IFASS) will allow direct entry of feature descriptors during manuscript preparation, eliminating the need to manually fill out, and then digitize feature analysis descriptor tables (FADTs). A further step toward automating the feature extraction process is the Computer Assisted Photo Interpretation (CAPI) Station. This improvement, scheduled for FY83, will automate film handling and will produce feature data directly in digital form, eliminating the need for LIS or AGDS processing.

Another improvement is the Extracted Feature Rectification and Processing System (EFRAPS). This system will facilitate monoscopic feature compilation by rectifying extracted features after compilation. This approach reduces the need for expensive stereocompilation equipment.

Even further automation will come about with the advent of the Digital Stereo Comparator/Compiler (DSCC). Its initial capabilities will include (at least) the ability to manipulate digital imagery of all types, and perform image

warping and stereocompilation. Having the imagery in digital form (instead of hardcopy, as with CAPI) will permit image enhancement techniques to be employed to facilitate semi-automatic feature extraction.

1.3 ORGANIZATION OF REPORT

The remainder of this report is organized as follows. In Chapter 2, black-and-white feature extraction techniques are reviewed and assessed. A concept of operation for a semi-automated feature extraction system is described in Chapter 3. Chapter 4 contains a discription of multi-spectral and multi-source techniques of potential use in further automating the feature extraction process. A corresponding concept of operation is described in Chapter 5. Finally, a summary of the important conclusions and recommendations of the FEAS is provided in Chapter 6.

2. REVIEW AND ASSESSMENT OF BLACK-AND-WHITE FEATURE EXTRACTION TECHNIQUES

This chapter reviews and assesses current technology for semi-automated feature extraction using black-and-white imagery. It is shown that the activities comprising feature extraction (e.g., search, detection, identification, classification and delineation) are independent when defined from a human viewpoint; however, it is not completely appropriate to define them as independent goals for a machine. The closest analogous activities for a machine might be "perceptual segmentation" and "recognition" of features, and these would most likely be interrelated, rather than independent, if machine perception were identical to human perception. Since it is not, machine perception goals must be tailored to interplay with human-oriented activities if they are to effectively semi-automate the feature extraction process. In this chapter, techniques are assessed in terms of their capability to achieve and satisfy the goals of machine perception.

The approach selected for assessing candidate feature extraction techniques is to analyze the general machine visual-perception problem in terms of an IU paradigm. This paradigm is a model for showing how the data within an image could be interpreted (in terms of its relationship to the physical properties of the objects it portrays) in order to permit the accurate inference of descriptive object properties. It also shows that the ability to solve complex recognition problems like feature identification hinges on the ability to infer such object properties reliably. The paradigm then shows that the only unambiguously-definable image properties from which

descriptive object properties may be inferred are global properties which can only be obtained under highly constrained (and, in fact, unrealistic) imaging conditions. The IU paradigm analysis therefore concludes that there are no image properties or characteristics yet identified from which specific object properties may conclusively and reliably be inferred. As a consequence, current machine perception technology cannot be expected to successfully solve the perceptual segmentation and recognition problems, nor be capable of fully automating feature delineation and identification.

The reviewed machine perception technology includes techniques characterized broadly either as computer vision or pattern recognition techniques. The focus of this review is on the two classes of techniques which are directed toward achieving the perceptual segmentation problem; i.e., edge extraction and region-based segmentation. Higher-level computer vision techniques as well as pattern recognition techniques, which in many cases assume that the latter techniques are available and reliable, are treated with less emphasis.

To reconcile the traditional view of edge extraction and region-based segmentation as fundamental low-level processes with the conclusion of the IU Paradigm, the rationale for the traditional view are examined and shown to be inconsistently defined and insufficiently related to physical properties. In particular, the notion of a "region" is especially ambiguous, and cannot be shown to be useful to either perceptual segmentation or recognition. On the other hand, while "edges" also are not clearly defined, the IU paradigm indicates that many feature boundaries do generate image structures which can be identified by edge extraction techniques.

The remainder of this chapter is organized as follows. Section 2.1 discusses our approach to assessing the candidate feature extraction techniques. Following a brief overview of the alternate evaluation approaches examined in Section 2.1.1, the evaluation methodology chosen - the Image Understanding (IU) paradigm - and its application to feature extraction assessment are discussed in Sections 2.1.2 and 2.1.3. Section 2.2 then reviews and assesses several major classes of feature extraction techniques in the context of the IU paradigm. The classes reviewed include edge extraction, segmentation, texture, statistical and syntactic pattern recognition, and symbolic techniques. Summaries of representative techniques contained within each class are also presented in Appendices A through E. Following the review and assessment of techniques presented in Section 2.2, the conclusions of the assessment and recommendations for future research are discussed in Section 2.3.

2.1 APPROACH TO TECHNIQUES ASSESSMENT

2.1.1 Overview of Candidate Approaches

This section discusses four approaches for assessing candidate feature extraction techniques. Following a brief description of each, the reasons for their elimination or selection are provided.

Analytical Evaluation - This approach to assessing the performance of candidate feature extraction techniques involves mathematically analyzing the operations each technique performs, mathematically characterizing the image data on which each technique operates, and predicting the performance of each technique accordingly. The low-level feature extraction techniques reviewed in Section 2.2 are in fact mathematical

operators, and respond predictably to mathematically-definable signal characteristics. However, there is little evidence at this time that the signal characteristics of DMA-relevant features as they appear in images are amenable to such characterizations. In other words, there is no evidence that every, or even any, type of feature has a unique or invariant "signature" from image to image, whether deterministic (like a known communications waveform detectable by a matched filter), or statistical (like a characteristic of some histogram which is a sample of a random process). This means that, currently, the performance of a technique in detecting, identifying, or delineating a feature cannot be predicted analytically with any confidence. Thus, evaluating techniques strictly by mathematical analysis is not a promising approach.

Literature Search and Inference - Another approach to technique evaluation is to examine published experimental evidence of machine perception technique performance available in open literature, and to extrapolate these experimental results to predict the performance of feature extraction techniques on representative aerial imagery.

There are, however, are a number of problems with this approach. One is that there are few results of applying algorithms to real imagery of any kind, let alone to aerial imagery. Many algorithms are based on an assumed set of mathematical characteristics or model of image behavior. Typically, the model is partly deterministic, to represent the signal characteristic of interest, and partly random, to represent noise (i.e., anything not of interest). Reported performance will then often be based on synthetic imagery, generated to ensure the original mathematical assumptions hold. These synthetic images usually bear little resemblance to real imagery; consequently, the synthetic results are not generalizable.

A second problem concerns the validity of the implicit assumption that if enough real-imagery results were available, meaningful performance measures could be predicted. Features are objects which are classified semantically, and different objects from the same class can vary significantly in appearance. Furthermore, images of the same object can be significantly different depending on imaging conditions. Thus, the results of how a machine perception algorithm performed on several images containing different instances of the same feature is not necessarily an indication of its performance for all instances of the feature.

Test and Evaluation - A third approach to technique evaluation is to implement all machine perception techniques of interest on a single system (for uniformity), and to compare their performance against a set of representative imagery. An example of this approach can be found in Laws (Ref. 1). An equivalent approach is to distribute a set of representative imagery to various research centers, where each center applies a different set of techniques directed toward the same extraction goal. A single group then assembles and compares the results from each center. This approach was taken in DMA's Pilot Digital Operations (Ref. 2).

One problem with these approaches is identical to a problem cited for the previous approach. The problem is that a representative set of imagery may be representative in the sense that it contains examples of all of the different types of features, but still it will contain only specific instances of each. There is no solid justification for generalizing an algorithm's performance on one feature instance to its performance for the class as a whole.

The more significant problem is that it is already known that virtually all techniques fall short of the performance required to meet DMA feature extraction production requirements. It is therefore questionable to rank-order the "goodness" of techniques based on experimental results, when most of the results do not satisfy required performance levels.

The Image Understanding Paradigm - A fourth approach is to review candidate feature extraction techniques in the context of the machine visual perception problem of feature extraction. The analysis framework underlying this approach is referred to as the IU paradigm, because it uses the same conceptual framework which is the basis of most current work in artificial intelligence approaches to vision and most of the research sponsored under the DARPA Image Understanding program. It is also similar to the methodology described in Kanade (Ref. 3).

The paradigm is applicable to any optical imaging problem for which the goal is to obtain higher-level information from the image. Higher-level information refers both to physical properties of the imaged objects and to the names of those objects. Obtaining the physical properties of an object may be regarded as a measurement process, while identifying its name may be viewed as a recognition process. Clearly, feature extraction qualifies as an Image Understanding problem. The recognition process is the problem of obtaining a feature's identity (e.g., FID or descriptor). The measurement process is the problem of locating the physical boundaries of a feature, and in some cases, mensuration for the purpose of obtaining a feature's height.

The IU paradigm shows how semantic categories relate to the physical characteristics of the objects to which they refer, and how semantic information is inferred on the basis of observed physical properties. It secondly shows how the latter physical properties (which have to do with objects and not images) are transformed to imagery, and what is subsequently required to infer object properties from the image.

Use of the IU paradigm serves the following three purposes with regard to reviewing the performance of machine perception techniques for feature extraction:

- It shows that feature extraction is a complex perceptual problem and helps place expectations about the performance of existing techniques in perspective.
- It shows that the signal nature of images which portray features is determined by object properties and the physics of the image formation process, and that there is no basis for assuming that the statistical or structural characterizations of signals (which do not explicitly model this process) ought to correspond to the meaningful object properties required to extract features.
- It describes the highly restrictive constraints which must hold in order to infer meaningful object properties from identifiable (and extractable) image characteristics.

2.1.2 The IU Paradigm

This section describes the IU paradigm which was chosen as the approach to evaluating the candidate feature extraction techniques. In the following sections, the ways in which information is represented and transformed within the paradigm are discussed.

Levels of Information Representation - There are three levels at which information exists or is represented in the IU paradigm (see Fig. 2.1-1). The "top" level is called the semantic level. Within the semantic level, physical objects are represented and referenced by names, codes or other symbology. Furthermore, it is in terms of such names that objects are described. For example, man-made objects which exist in human environments like home, office, and factor are named on the basis of their functionality. Natural objects such as flora or terrain are named according to natural properties which distinguish genus and variety.

The intermediate level is the physical level. It is at this level that real objects exist in three-dimensional space. The physical domain of an Image Understanding problem is the volume of space containing the objects being imaged. A physical domain is fully-specified if the visible properties

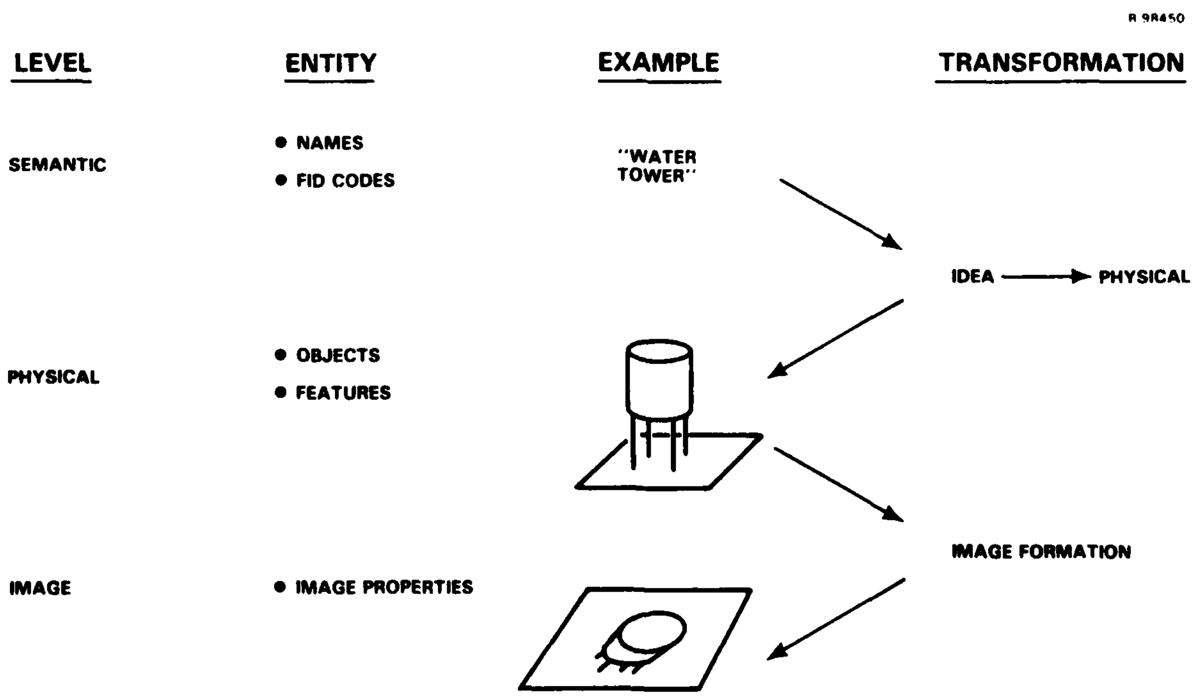


Figure 2.1-1 IU Paradigm Information Representation

of each object and the spatial relationships among all objects in the domain are themselves fully-specified. The objects within any particular physical domain of a given IU problem are "instances" of objects that could exist within the domain. Similarly, their spatial relationships in a given case are instances of how they could be configured. (This is simply to say that what a given scene contains is not completely known in advance).

The image level is the third and final level of the IU paradigm. This level consists of the data which represents the formed image of the physical world. Precisely how to characterize the image level and define the appropriate properties which contain significant information are not clear. Ideally, signal properties would be defined in terms of their relationship to the physical object properties which generated them. However, the information loss which occurs in the image formation process makes the relationship ambiguous.

IU Problem Goals - IU problem goals fall into the categories of recognizing objects (i.e., obtaining their semantic names) or measuring the physical properties of, or spatial relationships among, objects. Semantic names are about physical objects, and measurements are taken of physical properties of objects. In both cases, all needed information is contained in a complete description of physical object properties and spatial relationships. (Although to obtain semantic names, physical information must be related to semantic classification). How much information is needed will depend upon the particular IU problem goal and the possible variation among object properties and spatial relationships.

For the goals of object recognition and the measurement of properties of objects (rather than, for example, measurements

of distance between objects), the important information is visible object properties. Thus, it is important to know what such properties are, how they can be described, and what is required for their complete specification. There are hierarchical levels at which objects may be described. For example, an object may be decomposable into several distinct volumetric solids, like a table top and its legs, and then the relationship among these components parts can be specified. The following discussion will focus only on visual surface properties of object, the most fundamental level.

The major categories of visual properties are shape, color (or albedo), reflectivity, and size. Visual properties are characteristics within each category. For example, color properties include "red" and "blue". Shape properties include "cylindrical" and "oblong." Each category is describable by local or global properties. Local properties are quantitative and specify a category's value precisely at a small area or patch on an object surface. If the local property is specified at every small path on the object surface, then that visual category will be fully determined for the entire object.

Of all the categories of visual properties (and in particular for our problem), shape is the single most important for recognition. If the shape of an object is completely known, then that information is usually sufficient for recognition. Knowledge of an object's color or approximate size may provide clues for recognition, but are generally not sufficient for recognition except for highly constrainted problems.

The shape of an object surface is precisely defined when, with the object located in some spatial coordinate system, the locus of coordinate points which define its surface can be

specified. Size is precisely specified when the coordinates are given in terms of dimensional units of distance. Thus, one way of completely specifying surface shape is to define the object surface in a Cartesian spatial coordinate system, and specify the elevation of the surface as the value of the function at each (x,y) coordinate. Surface elevation is not actually a local shape specification, but clearly surface shape is fully specified in this way.

Local shape specifications are essentially first or second derivatives of the surface elevation function. The orientation of a small patch on a surface is analogous to the slope at a point on a function of one variable. Orientation of a patch is specified by two parameters. If the orientation at each patch on a surface is specified, the entire surface function (minus a constant term) can be recovered, just as an original function can be recovered by integration from its derivative, or slope, function. Thus, local specification of orientation also determines surface shape.

There are various equivalent methods (see Fig. 2.1-2) of specifying orientation, which is always given with respect to some coordinate system. It may be specified by two directional derivatives, by two of the three possible direction cosines of the normal vector to the surface patch's tangent plane, or by the gradient magnitude and direction. The important point is that all of these equivalent representations require two parameters. Thus, local shape has two degrees of freedom and requires two pieces of information to be fully specified.

Global shape properties describe the overall shape of an object. Qualitative properties like "box-like," "egg-shaped," or cylindrical are used in natural language to describe shapes generally. Quantitative descriptions give a precise specification

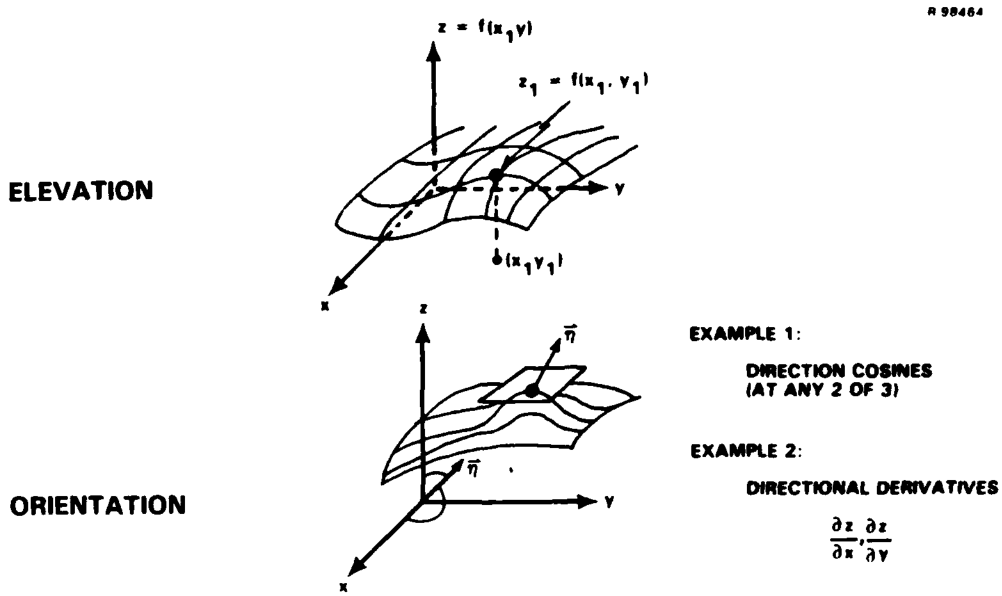


Figure 2.1-2 Methods of Specifying Orientation

of shape. Such descriptions may consist of individual words like spherical or cubical which specify shape unambiguously, or parametric forms like the equation of an ellipsoid with its parameters specified. Characteristic global descriptions like "polyhderal" or "quadric" specify certain constraints on shape, but are not complete specifications.

<u>Semantic - Physical Level Transformation</u> - One way of regarding the relationship between the semantic and physical levels of information representation is as a transformation from the named concept of an object to a particular physical instance of that object in space. The descriptive entity at the semantic level is the concept, while the descriptive entities at the physical level are the various physical properties which describe the object's appearance. Thus, the transformation is from concept (referred to by semantic name) to physical property set (which describes the object's appearance). Note, however, that the physical property set must be sufficient to

describe all variations in appearance for all possible object instances of the semantic concept or semantic class. All possible object instances means all semantic class instances possible for the IU problem under consideration.

The more generic the semantic class, the greater will be the variation of properties in the property set. For example, the set of all chairs exhibits greater variation in appearance than does the set of all armchairs. A given IU problem will have a finite set of semantic classes which must be recognized if one of their member objects is initiated in the scene. Which classes are required, and how specific or generic is the nature of each class, depends on the goals of the IU problem.

Recovery of Semantic Information - The previous section stated that a semantic concept can be related to the physical appearance of its possible object instances by regarding the concept name as the name of a set containing physical properties sufficient to describe all such instances. To be more complete, it must be mentioned that the characteristic set of physical properties is actually a set of property values, or value ranges within the bounds of which the property values of all member objects must fall. To describe the members of a single semantic class, it may suffice to select properties and value ranges which describe only the rough appearance of those objects with sufficient detail to describe their variations with respect to one another.

However, in the recognition process, the reverse transformation, or association, must be performed. That is to say, given a sufficient set of physical properties and values which describe an object appearance, the problem is to recover the

semantic classification, or name, of that object. What constitutes a sufficient set of physical properties will depend on how similar other objects in the environment outside the given semantic class are to any object instance of that class. As long as no other object is identical in appearance to one of the objects, there will exist a "unique property set." This set will permit any object whose physical properties fit the unique property set to be properly identified as instances of that semantic class.

If, however, an out-class object is arbitrarily similar to some in-class object, the unique property set may be specified in terms of quantitative local properties. As discussed earlier, the sum of all local properties completely specifies the physical appearance of an object. Thus, as long as all in-class objects are distinct from all out-class objects, the unique property set may at least be specified by quantitative local properties. <u>In order for identification on the basis of visible appearance to be possible, there must exist a unique property set for every semantic class</u>.

None of this discussion defines which properties are best to use for identifying unique property sets for a given IU problem, nor how to go about choosing them. Furthermore, what kinds of properties can or should be recovered from an image has not been identified, nor is it known what properties the human vision system uses as the basis for recognition. However, the capability of human vision is proof that for very sophisticated recognition problems, including feature extraction, there not only are unique property sets for each recognizable class of objects, but that these sets are recoverable from images.

In summary, what physical properties are needed to perform recognition depends on what properties are sufficient

to distinguish object members of one semantic class from all possible out-class members in the environment. The fact that there can be considerable variation of properties among objects within a class, and similarities between the properties of objects across different classes, contributes to the complexity of the recognition process. Regardless of how complex the recognition problem, if objects are distinguishable, their distinctive appearance is derivable from local physical surface properties.

Physical-Image Level Transformation - The previous section emphasized that sets of physical properties required for recognition are based on the value of local properties at each point in an object surface. The transformation from the physical level to the signal level is achieved by the image formation process. An in-depth discussion of the physics of the image formation process can be found in Horn (Refs. 39 and 40). Images of objects are also formed on a fundamentally local basis, and in this section it will be shown that this process involves an inevitable loss of the characteristic local object surface information. This loss is so severe that, for the conditions under which images are formed for subsequent feature extraction, an image pixel intensity value alone conveys no information about the local surface properties which played a role in its determination. Thus, recovering any kind of physical property information will have to be based on global image properties, i.e., on the behavior of collections of image pixels.

Each image plane path of smallest resolution corresponds to a ray which extends outward from the camera to the scene and intersects a small patch on the first object surface encountered. For the purpose of this discussion, the patch is assumed to be of infinitesimal area, so that its orientation

is defined by the direction of a vector normal to its tangent plane. If a point light source is assumed, the proportion of incident light reflected from the infinitesimal patch in the direction of the viewer is determined by the reflectivity function of the surface and by the three angles shown in Fig. 2.1-3. The three angles are:

- i, the angle of incidence with respect to surface normal
- e, the angle of reflectance with respect to surface normal
- g, the angle between the incident and reflected rays.

If there are multiple light sources, the total reflected light in any direction is the sum or integral of the reflected light contributed by each infinitesimal point source.

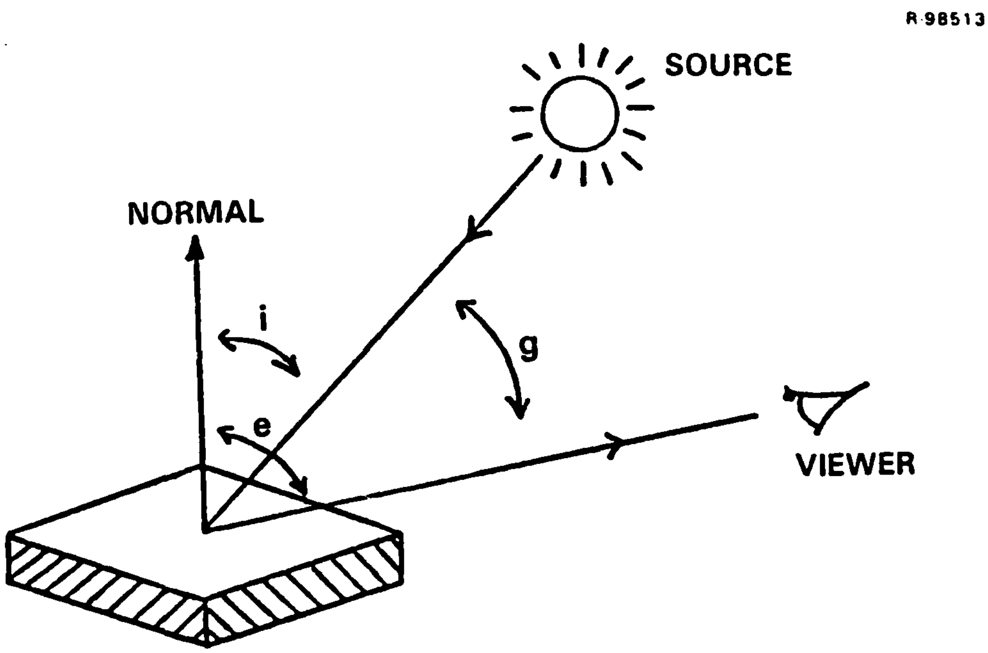


Figure 2.1-3 Image Formation Process

The physical laws which relate the local image property of irradiance to local object surface properties via the imaging conditions also govern the way in which global properties of objects will be transformed to global properties of images. The difficulty is that orientation, the local specification of object shape, is specified by two degrees of freedom, whereas image gray tone is specified by only one parameter. Thus, even in constrained conditions where surface albedo and reflectance are known, as well as illumination and viewing angles, information loss will occur. Each pixel value is determined as a function of the properties of a local object surface patch and the imaging conditions, independent of the properties of adjacent object patches. Thus, local shape information, which was shown to determine global shape information, gets lost in the image formation process.

<u>Recovery of Physical Information</u> - Despite the loss of local shape information, certain global object properties can be inferred from indentifiable image properties if sufficient constraints are in place. To demonstrate this, assume that the imaging conditions are constrained by placing the illuminant a sufficiently long distance from the object being imaged. Also assume that the viewer is located far from the object surface. These conditions clearly hold for aerial imagery analysis problems. If the visible object surface is planar, then the angles of the incident and the reflected light rays (with respect to surface orientation) will each be constant at every surface point. If the reflectance is constant across the surface, the same amount of light will be reflected from each point. If the albedo is constant, the same frequency spectrum will be reflected from each point. Assuming the image sensor to be ideal, the planar object surface will appear as an image region of constant gray-value whose shape is the geometric projection of the planar object

surface. In cases for which the imaging constraints hold, for which the object surface is planar with constant reflectance and albedo, and for which the spatial relationships are such that the object will not be obscured, then the image property, adjacent-patch-of-pixels-with-constant-gray-level, will be a valid image property which corresponds to a global object property, planar-surface-of-constant-reflectance-and-albedo. Its value could be defined as its gray-level, or as some description of its boundary shape.

To demonstrate how restrictive these constraints are, consider how relaxing certain of them eliminates the image patch property of constant gray level. If all constraints hold except that the surface is not planar, then the image patch pixel gray-levels (except for special reflectance functions) will vary as local object orientation varies. If all constraints hold but albedo is not constant, i.e., the "color" of the surface changes, the gray shade will change correspondingly. For these cases, gradual changes in local object property may cause gradual changes in gray value, so that the image region may be nearly constant. However, if the reflectance function is specular, a gray-level discontinuity can result, completely spoiling the image property.

2.1.3 Application of the Paradigm

This section demonstrates how the feature extraction problem fits into the framework of the IU paradigm. The demonstration begins with a discussion of how features can be categorized according to complexity of recognition, and a simple class of features is identified which is consistent with the IU paradigm discussion of object recognition by unique property sets. Next, the aerial image formation process is examined and found to have insufficient constraints; therefore, direct recovery of physical properties is not possible.

With feature extraction thus characterized as an IU problem, the delineation goal of feature extraction is discussed in terms of the paradigm, and segmentation processes are shown to be required for automatic delineation. Next, the notion of edges and regions as meaningful signal properties is discussed, and it is suggested that the most intuitively appealing basis for their validity exists independently of whether there is reason to believe that object properties can be inferred from these image properties. This section concludes with the observation that many object properties do generate image edges, or lines, but that homogeneous regions found by current segmentation algorithms seem to be generally unrelated to object properties meaningful within the feature extraction process.

Feature Classes - This section describes certain broad categories of features according to physical appearance or configuration in order to lend concreteness to the analysis of the feature extraction problem. The reference list of features for this discussion is the DLMS Level V Specification (Ref. 4). This section also identifies a major class of features which are recognizable and are similar to the kinds of objects discussed for the IU paradigm.

The physical objects which comprise features can be divided topologically into either surface objects or volumetric objects. A surface object is any feature which is part of or lies on the earth's surface and follows its elevation contour. Surface objects may be man-made, natural, or cultivated, and include natural land-based features like croplands and ground-surface categories, all bodies of water, as well as many cultural features like roads, railroad tracks, and athletic fields Volumetric objects protrude from the earth's surface, and include vertical obstructions, buildings, towers, and trees.

Features may also be categorized as either individual surface or volumetric objects, or as composites of these types. Individual features are volumetric or surface types which qualify as features independently of whether they are also part of some composite type. Individual volumetric features are single structures, like offshore platforms, (single) buildings, smokestacks, billboards, and windmills. Examples of individual surface features are grasslands, ice areas, and aqueducts. Composites are defined as features which exist by virtue of the sum of their parts, where the parts are either qualified individual features or spatially non-adjacent objects which are not on the feature list. Individual features may overlap in space without being considered composite. For example, a tower may be located in the middle of a meadow, but if each qualifies as an individual feature independently of the other, it is not considered a composite. Examples of composites which may be composed of individual features are power plants and refineries, which may contain smokestacks and conveyers, both individual features in their own right. Composites which may not contain other (DLMS) features, but which are comprised of various objects not on the feature list, include schools, hospitals, and prisons.

Many individual features require context for recognition, so individual features in general is not the class which corresponds directly to these discussed in the IU Paradigm. While it is not immediately clear which features do and which do not require context, an example can be given to illustrate the idea. Many single-family residences would not be readily-identifiable as such if the analyst were not using, whether to his conscious knowledge or not, contextual information like:

- This building appears in a region which is a residential area

- Immediately adjacent to this building is a two-car driveway and a small grassy area (a lawn).

If a mask were placed over the entire image, with a window the precise size of the image projection of the building, permitting only the building to show, and with no other a priori or collateral knowledge available, the analyst would probably quickly identify that the image patch portrayed a building, but he could well have difficulty making the conclusion that it is a single-family residence. Whether he could or not, and how quickly, would of course depend on the effective imagery resolution, how distinctively house-like the building was, etc. The point is clear, nevertheless, that context plays an important role in the identification of many features.

Thus, it is the class of non-contextual (individual) features which are similar to IU paradigm objects. Some examples of these are volumetric features which have distinctive shapes or are only generically classified, e.g.,

- many "associated structures" (DLMS FID 1800's) in the Industry category, like building (generic), smokestack, rotating crane on tower, cooling tower, powershovel
- certain recreational and amusement-park features (FID 3000's) from the Commercial/ Residential category, like domed stadium, grandstand, rollercoaster
- various types of towers (FID 5000's)
- certain distinctive governmental and institutional buildings (FID 6000's), like house of religious worship with temple, arch, pyramid
- storage facilities (FID 8000's) like cylindrical tank, water tower, silo.

Many individual surface-type features are most likely identifiable by virtue of their shape, or, for the case of natural features, by their "textured" appearance. Examples are:

- lineal features in the transportation category (FID 2000's): railroad tracks, canal, roads, bridges, aqueduct
- athletic field
- certain Military/Civil installation features (FID 7000's): runways and taxiways, breakwater/jetty, wharf/pier
- most landform and vegetation features.

Aerial Image Formation - The nature of aerial imagery generates both simplifying constraints and added complexity with respect to earth-based Image Understanding problems, like understanding images of indoor office environments or outdoor urban environments. The major simplifying constraint is that there is only a single source of illumination, the sun, that is very nearly a point source, and that both the sun and the viewer are generally optically distant. This means that the angles of ray incidence and reflection over relatively large local areas of the image depend only on object surface orientation. If an object surface has constant reflectance and albedo, the gray-level variations in the corresponding image patch will depend only on surface shape.

A second simplifying constraint is that of feature topology; namely, there are surface features which are essentially two-dimensional. If they were perfectly two-dimensional (i.e., planar surfaces) and assuming no perspective distortion, the only image shape distortion would result from relative tilt between the image plane and the object surface. This creates a "foreshortening" distortion, in which the distortion

is contraction of distance in the direction of tilt. For surface features on undulating surfaces, shape distortion will be minimal if the view is taken directly from above. In any event, the image shape of surface features is much more representative of their physical shape than is the image shape of volumetric features in general, due to the numerous possible silhouette shapes with different orientations with respect to the image plane. Whether the shape of surface features is useful for recognition depends on whether the shape can be recovered from an image, and whether the shape is a distinctive physical property for the class.

A third simplifying constraint unfortunately is not a dependable one. It is the fact that spatial relationships among features are much less three-dimensional than for earth-bound problems in which the three-dimensionality of space and objects surround the viewer. In aerial imagery, all objects are located on a single surface, and so are constrained to lie in or on this surface. If projective imaging geometry were purely orthographic, and taken directly from above, there would be no occlusions. However, stereo imagery requires that images not be taken vertically, both in order to achieve parallax for stereo fusion and in order to obtain an acceptable side-view of vertical objects for the purpose of monoscopic height mensuration. This means that occlusion of other objects will occur, and an unobscured view of surface-type feature shape cannot be guaranteed. Furthermore, shadows will likely exist, spoiling the opportunity for high-contrast boundaries or homogeneous surfaces for many features.

Despite the simplifying constraints described above, aerial imagery presents some overwhelming difficulties for the capabilities of current machine perception technology. These

difficulties are generally due to level of detail. Aerial imagery contains enormous detail due to the macroscopic scale of what is being imaged. This detail prevents object surfaces from having homogeneous properties which would be desirable for existing machine perception techniques. Whatever the resolution of the imagery, many macroscopic objects get averaged into the irradiance of a single pixel. Rarely are the objects patterned in a regular enough fashion that the effect is to create a homogeneous patch on the image.

Detail phenomena can be divided into two categories: detail which is part of a surface itself, and detail which results from other (small) objects resting on a surface. An example of the former is roof surfaces. In a typical aerial photograph, roof surfaces (perhaps as a result of weather-beating) exhibit non-uniform splotches and discolorations. Upon inspection, the roof may appear to be more or less homogeneous with respect to the surrounding imagery, where in actuality the observor's visual perception system has done such a good job of identifying distinct objects that they appear to be homogeneous after the fact. Another example is vegetation areas. A grassland area, like a meadow, may be quite homogeneous-looking compared to the adjacent forest and residential areas, but its homogeneity can more often than not be upset by stray shrubbery, changes in types of grass, etc.

The second source of detail is small (with respect to imagery scale) objects which rest on top of features. A good example is vehicles which cover roads and highways. Even if roads were never obscured by buildings and trees, there would still be the problem that rows of parallel-parked cars or heavy traffic in right-hand lanes spoil the nice edge-like behavior of their boundaries. Roofs also suffer from this kind of detail.

Roofs tend to be dotted or covered with such objects as ventilation outlets, air-conditions systems, roofdecks, skylights, etc.

Goals of Feature Extraction - The IU paradigm focused primarily on the recognition problem. It emphasized that if in-class objects are discriminable from out-class objects, there must exist "unique property sets" describing those objects. It also showed how quantitative local properties (which provide the basis for global descriptions) get unavoidably lost in the image formation process, and that highly-restrictive constraints must hold to infer global object properties from identifiable global image properties. The implication was always that the purpose of recovering such properties was for feature identification. However, the other perceptual feature extraction goal is feature delineation, an even more time-consuming process for human analysts.

One way of regarding perceptual delineation is as a process of physical property recovery, just as for identification, but now the physical properties are the spatial position of the feature on the earth's surface. A better way is to regard delineation as a segmentation process. Segmentation as used here does not mean dividing an image into regions with homogeneous signal properties, as do segmentation algorithms. Rather, it means the perceptual process of distinguishing an object (in this case feature) from its background. More specifically, it is the process of distinguishing which pixel patches on an image correspond to projections of the same object entity. Features are by definition always such that the image projection of a feature's visible surface will always correspond to a corrected image patch (i.e., a patch whose boundary forms a single closed contour). In human perception, it is not known whether such figure-ground organizational phenomena occur without invocation of world knowledge or not.

In other words, the segmentation process may be intimately tied to the recognition process. In any case, recognition certainly cannot proceed before segmentation has begun. A set of object properties could not be matched to a unique property set until it were known that the properties all corresponded to a single object.

Thus, delineation is based on the perceptual process of segmentation which recognition may not precede. The next sections discuss the notion of edges and regions, often regarded as fundamental primitives of images, and how these relate to visible feature boundaries as they are initiated in images.

The Notion of Edges and Regions - Edges have been traditionally regarded as fundamental image primitives (Ref. 7). There are at least two strong bases on which to argue that edges are fundamental image primitives, and therefore, that edge extraction ought to be used as a perceptual segmentation process. One is that etchings or line drawings of natural scenes convey an enormous amount of information about scene content. While this is clearly true, the fact is that there are many clues besides object boundaries that enable humans to recognize the contents of a natural scene. To trace out object boundaries in a natural photograph, a subject would use this other information to show him where an otherwise faint or non-existent boundary would be located. These clues would also provide him the information needed to properly ignore intensity discontinuities that did not correspond to genuine object boundaries.

Another argument for the importance of edges is based on an information-theoretic view of an image as a two-dimensional signal. To drastically oversimplify the ideas of information theory, signal events which occur most infrequently are the ones with the highest information content. If it is assumed

that most images consist of adjacent regions of near-constant intensity, then the spatial shape and the step size of these boundaries would be representative of the image information content. The boundaries are the image edges, and their duals are the image regions. According to this model, either the edges or the regions are sufficient descriptions of the image, and are in this sense fundamental image properties.

Edges, Regions, and Feature Properties - The previous section discussed rationale for regarding edges and regions as fundamental properties of images. This section discusses whether such properties correspond to meaningful object properties. The discussion begins with the issue of the relevance of the properties themselves, and then proceeds to address whether such properties are identifiable or extractable by machine perception algorithms. The section concludes with a discussion of how such properties and algorithms may be applicable to feature identification or delineation problems.

It is clear that image edges are often generated by meaningful object properties. An image edge is generated by an abrupt change in image intensity along a continuous contour across part of the image. There are a number of object phenomena which can cause this. One is a discontinuity of surface orientation, (e.g., when two planar surfaces intersect). If their color and reflectivities are constant and identical, then the orientation differences will cause each plane to be a different gray shade, forming an edge at the intersection. An example of this is the corner of a building.

Another source of image edges is occluding boundaries. The visible surface of the occluding object will project to the image as a region with approximately one gray-level, and since the visible part of the occluded surface is unlikely to

have the same color, reflectivity, or orientation, it appears as an adjacent image patch of different gray level. Thus, occluding surfaces will generally be three-dimensional object-type features rather than surface type. Surface features, like roads, will usually have boundaries which mark a change in surface material (and thus, in either surface color or reflectance). Such boundaries also create image edges. Finally, edges may be created by illumination boundaries, which for aerial imagery means shadow boundaries.

There are unfortunately a number of difficulties associated with interpreting image edges as meaningful object properties. One is simply the fact that there are a number of properties which generate edges, so that determining the cause of any particular edge is difficult. Another is that not all edge-generating properties, like the boundaries of features, always generate edges. A third is that image edges are not well-defined. Some edges are fuzzy and correspond to slowly-varying contrast changes, while for others, the contrast varies significantly along the extent of the contour. This lack of definition helps to explain why some feature boundaries do not seem to generate edges. The fact that a human can see faint boundaries of objects may be due to after-the-fact extrapolation of a known object's boundary. This lack of definition means that edge extractors will find some boundaries and miss others. Moreover, they will find edges that do not really exist, but are caused by specularities or unwanted surface detail.

It is much less clear what kinds of physical phenomena generate distinctive image regions, or whether such properties are useful for machine perception at all. Detail phenomena tend to ruin the potential homogeneity of many kinds of surfaces. In addition, although object or feature image patches may look homogeneous to an observer it does not imply that they are homogeneous according to some signal measure.

The primary difficulty with regions is the necessity of defining them in terms of average properties of pixel intensity. It was shown that for all pixels of a region to be of identical gray-level, unrealistically-restrictive constraints had to hold. Even if all constraints held except the ideal sensor constraint so that a few levels of noise would be added to the constant gray patch there is still the difficulty of how to extract, or identify, such regions (i.e., the segmentation problem). In other words, even if images were composed of a partition of tiled constant gray level regions, with additive noise in each, it is not clear that any segmentation technique would identify each region, without strong constraints on:

- the minimum size of each patch
- the maximum level of additive noise
- the minimum difference between average gray-level of adjacent regions
- the shape of each region.

These constraints are necessary, since the averaging process requires a minimum number of pixels to obtain meaningful statistics, and pixels near the border of a region would easily be noisy members of the adjacent region.

Perceptual segmentation was defined to be identification of closed boundaries (or interior regions) of projected feature image patches. Perceptual segmentation is equivalent to the problem of feature delineation. Region-based segmentation algorithms, on the basis of the discussion above, do not seem useful in principle for performing automatic feature delineation. Moreover, experimental work supports this conclusion (Ref. 2).

Edge extraction seems a more promising approach to the perceptual segmentation problem, since many physical feature boundaries do generate image edges. That edge extraction would be more promising than region-finding for perceptual segmentation is not surprising since edges are more "local" than regions. Edges are image properties which occur essentially along a simple curved region of an image, whereas regions must maintain a definable signal property over greater image extent in order to be useful. Of course, since image edges are not well-defined either of themselves or in their relationship to feature properties, edge extraction techniques cannot fully automate the feature delineation problem either. How such techniques may be used, however, to semi-automate the problem is discussed in Chapter 3.

2.2 REVIEW AND ASSESSMENT OF FEATURE EXTRACTION TECHNIQUES

2.2.1 Overview of Techniques Reviewed

Automated techniques for feature extraction can refer to many varieties of computer-related technologies, depending upon what phase or activity of the overall feature extraction process is the target for automation. The target activities for technique assessment in the FEAS are feature identification and delineation, both of which require visual perception. The assessed techniques will therefore be referred to as machine perception techniques. The primary goal of visual machine perception is to analyze an image, pick out and identify objects, estimate their positions in space, and thus to see as humans do. Approaches to visual machine perception have originated from a variety of fields, including electrical engineering,

computer science, statistics, and psychology. The state-of-the-art is in some disarray, with machine perception papers published in a variety of journals with a variety of motivations and goals.

The techniques reviewed can be divided into two general categories (Table 2.2-1): computer vision and pattern recognition techniques. Pattern recognition has traditionally been recognized as a field in its own right. Its mathematical basis lies in statistical decision theory, and is based on the assumption that various patterns which one wants to classify, or among which one wants to discriminate, can be uniquely represented by the statistical behavior of measurements (or "features") of those patterns.

TABLE 2.2-1
TECHNIQUES REVIEWED

R-98519

		REVIEW EMPHASIS	ALSO TREATED
"COMPUTER VISION"	EDGE EXTRACTION, SEGMENTATION	√	
	TEXTURE	√	
	SYMBOLIC DESCRIPTION, MODELS		√
	SYMBOLIC MATCHING		√
"PATTERN RECOGNITION" TECHNIQUES	STATISTICAL		√
	SYNTACTIC		√

More recently, the idea of representing the structural aspects of patterns as grammatical constructs of pattern primitives has received attention. This approach is usually called syntactic pattern recognition. It is a method of analyzing and matching pattern shape or structure, and assumes that the essential nature of patterns is deterministic rather than random as does statistical pattern recognition. Pattern recognition techniques *per se* are surveyed briefly in this report, but are not analyzed in detail since they do not address the perceptual segmentation problem of low-level vision described further below.

The remaining machine perception techniques can be characterized as computer vision techniques. To some extent this characterization is implicit acknowledgement of the fact that it is misleading to view real-world objects (as they appear in images) as being representable by patterns. Pattern connotes a limited, two-dimensional structure which does not seem powerful enough to represent the variety of forms exhibitied by image-instances of a (semantic) object class. Pattern recognition techniques normally involve only a single-level representation of an image as the basis for recognition. That is, pattern recognition techniques extract primitives, or measure features of an image and compare them to paradigm or model representations of prototypical patterns for the purpose of classification. The computer vision view of object recognition generally builds up several hierarchical representations of an image, and makes recognition decisions at the higher levels (Ref. 6).

The focus of this review will be computer vision techniques which perform extraction at the lowest level of the hierarchy. The two major classes of low-level techniques are edge extraction and region-level segmentation. These low-level

vision techniques are so-called because they operate on pixel intensity values directly. One goal of low-level vision techniques is to group pixels into regions which correspond to single objects. This is the perceptual segmentation problem discussed earlier.

The perceptual segmentation problem has proven to be exceptionally difficult for machine perception. To date, there are no segmentation algorithms which can completely partition a complex, real-world image into distinct connected regions, so that each region corresponds to a single object's surface, and so that each such surface is represented by only one region. This state of complete partition is the goal of region-based segmentation (or simply segmentation) algorithms.

Even the less-ambitious goal of finding image edges, *each of which corresponds to some physically meaningful boundary*, is extremely difficult, and explains the proliferation of edge extraction algorithms. Edge extraction algorithms pursue a less ambitious perceptual segmentation goal in that they do not attempt to obtain a complete partition. Finding meaningful edges is nonetheless difficult, because it requires making global perceptual decisions based on the ambiguous local evidence of detected edge elements. This problem is discussed in greater detail in the review below.

Textures also are often regarded as a low-level vision technique class, since they do measure and represent the image directly. However, except for synthetically-generated imagery, texture measures are not perceptual segmentation processes. Simple texture measures may, however, be embedded in region-based segmentation algorithms.

Higher-level computer vision algorithms will also be surveyed briefly. These include computational representations for two- and three-dimensional shape, as well as matching techniques for comparing stored object representations to extracted representations. In general, high-level techniques suffer from the fact that existing low-level techniques do not perform satisfactorily. For this reason, high-level techniques are only reviewed briefly.

Recent surveys of computer vision techniques may also be found in Refs. 5, 7 and 8, and with more an image-processing flavor, (Ref. 9). Standard references in statistical pattern recognition are Refs. 10 and 11.

2.2.2 Edge Extraction

This section describes a conceptual model of edges in order to motivate a discussion of the edge extraction process. The basic phases of the edge extraction process are then presented, with discussion of alternative methods for accomplishing each phase. Finally, the problem of how to assess such techniques is discussed, and a recommendation is made in favor of a general-purpose edge extraction algorithm for semi-automated feature extraction.

A Model of Edges - As a means of understanding edges and edge extraction algorithms, it is useful to develop a conceptual model of image edges. In this model, the image is viewed as a continuous (not digital) intensity function of two variables. If one defines the gradient image to be a separate function whose value at each coordinate point is the gradient magnitude of the image intensity function, then edges are defined as continuous contours which are projections of gradient image ridges onto the plane defined by the x and y coordinate

axes. A ridge is composed of a locus of points which forms a connected curve in space. Each point on a gradient image ridge is a point of local gradient magnitude maximum in the gradient direction. (In order words, a cross-section of the ridge would show the ridge point to be a local maximum of the other cross-sectional points). Note that the gradient direction at a ridge point is always perpendicular to the direction of the ridge contour at that point.

This continuous model is, of course, only a conceptual one. It portrays edges functionally as having a continuous range of values (or intensities) and a continuous domain of (x,y) coordinate values. Digital images, however, are discrete both in intensity and pixel size. This discrete representation causes substantial difficulties in attempting to define edges for digital images. If intensity contrast along an edge contour changes too quickly with respect to pixel sampling rate, the contour will not seem smooth due to the staircase effect of pixel values at the boundary. If the intensity changes quickly in the gradient direction (i.e., perpendicular to the contour), the pixels at the boundary will average portions of both sides of the edge, also causing smoothness to be lost. The fact that, for aerial imagery in particular, very few regions have constant reflectance, albedo, or orientation means that the underlying continuous boundaries will not be smooth, and the quantization effects described above will be prevalent.

In spite of these and other difficulties, the continuous edge model is the model on which most digital edge extractors are conceptually based. They attempt to estimate the gradient image in the edge detection phase, to find which points are points of cross-sectional maxima in the thinning phase, and to link them together according to expected contour shape in the linking phase.

Edge Detection - This phase is sometimes called gradient estimation, since the purpose is to measure how fast and in which direction the image intensity function is changing at a given point, which is what the gradient magnitude and direction functions show. Estimation refers to the fact that the image is discrete, and measurement of the derivative function must be approximated by taking differences.

Typically, a series of difference operators is applied to each point in the image. Each operator measures the extent to which the intensity is changing in a particular direction defined by the operator. Gradient estimators, like the Roberts (Ref. 12) and Sobel (Ref. 13) operators, estimate directional derivatives in two perpendicular directions, and then compute gradient direction and magnitude from these. The compass gradient method (Ref. 14) applies several operators to each point, each of which is an edge template which measures the strength of the directional derivative in a unique direction. The operator with the highest output is chosen as the direction of change, with its output level used as the magnitude.

Gradient estimators are first-derivative intensity operations. Also used are second-derivative operators like (Refs. 15, 16). If edges follow contours of local gradient maxima, they can also be said to follow contours of second-derivative function zero-crossings. The Marr-Hildreth operator combines bandpass filtering or smoothing with a rotationally-symmetric second-derivative Laplacian operator.

Another approach is to detect edges of various widths, by using edge operators of multiple sizes. The justification for this approach is that edges exhibit a variety of widths, and it is best to have different sized operators, each of which is most sensitive to edge changes at a given resolution. The

size of an operator should be roughly the same as the edge width. If it is too large, the edge will get averaged out. If it is too small, the edge may look like a homogeneous region. In the spatial-frequency domain, the notion is that information in an image is distributed throughout a two-dimensional frequency spectrum, and that bandpass filtering will isolate which information is significant in a particular range. Examples of edge detectors which employ multi-resolution edge operators can be found in (Refs. 15, 17, 18).

Thinning and Thresholding - Second-order operators produce zero-crossings at points of local gradient maxima, and zero-crossings provide a readily-identifiable indication of such locations. With a gradient function itself, however, local maxima can only be determined by comparison with other local values. Thinning and thresholding are two operations intended to identify the significant points in the gradient image.

Thresholding is a simplistic method of finding gradient maxima, and does not work well for complex imagery since it is based on the assumption that significant gradient values will always be above a particular value. Whether the threshold is chosen in advance, or selected as a point between two peaks in a histogram, the fact is that many physically-significant edges will have low-contrast image boundaries and thresholding will eliminate them.

Thinning is a more judicious means of identifying significant points. Thinning is based on the fact that application of differencing operators across an intensity contour (i.e., in the direction of the gradient, or perpendicular to the direction of the edge) will produce a series of outputs, the center of which should be a maximum and correspond to the

center of the edge. Thus, thinning algorithms look for the central thread of pixels which form the ridge of the gradient. However, they still operate on a local basis, making decisions based only on a few nearest neighbors. A typical thinning algorithm is discussed in Ref. 19.

Much about the thinning process is necessarily arbitrary. For example, thresholds must be chosen to decide if an element is sufficiently aligned directionally with its neighbors to be considered part of an edge, rather than as noise. The more conservative a thinning algorithm is in its decisions not to eliminate edge elements, the less will be its likelihood of eliminating valid but low-contrast edge elements and the greater will be its tendency to retain noisy elements.

Linking - Once edge elements have been detected and thinned, the difficult task of linking or grouping them into meaningful boundaries remains. This phase is more difficult than earlier phases because it is a perceptual organization process as described in the technique overview. This requires a higher-level integration process not required of edge detection or thinning. Edge detection is essentially a local, non-interpretive process. The strongest decision made by edge detectors is to choose what direction and what magnitude of intensity change a group of pixels seems to be indicating. This is more nearly a signal processing operation than a perceptual operation. Deciding which local edge elements belong to the same more global entity is a process of perceptual organization.

What makes the linking process so difficult is the discrete nature of the image and the related problem that edges which are meaningful (by virture of their relationships to object properties) are not well-defined. If images were continuous

functions like the model described above and continuous gradient-images were available, whether or not a gradient point was on a ridge could be easily determined by examining arbitrarily-close points. In discrete images, gradient magnitude and direction are necessarily quantized. Even if an edge is relatively simple, (in the sense that it is relatively straight and has high, even contrast along its length) it may be oriented unfavorably on the rectangular image pixel grid, so that neighboring edge-elements do not have uniform magnitude and direction.

The linking process becomes simplified if the possible shapes permitted edges are constrained by higher-level knowledge about what kind of edges will be present. If edges are known to be of polynomial form, the Hough transform (Ref. 20) may be used. The most commonly-used polynomial is of course the straight line, and this is the type of edge which appears frequently in urban areas of aerial imagery.

Linking procedures which do not rely on a specific global shape must use some sort of search procedure. Such procedures start with a given edge element chosen, for example, according to high gradient magnitude. The procedure then examines neighboring elements to determine whether they meet criteria for inclusion in the link. Criteria are based on some measure of gradient magnitude and/or directional continuity over the length of the link. The criterion may be defined very locally and based only on neighbor elements, or less-locally over several elements, in which case constraints on curvature and continuity of average contrast may be used.

<u>Assessment</u> - The function of an edge extractor should be to find and represent edges in an image, where the representation provides sufficient information to reproduce the original edge, both in terms of its shape, image orientation and

position, with sufficient accuracy that the desired degree of original detail is preserved. Edge extractors will therefore be assessed according to their ability to achieve this goal efficiently. This result of such an assessment can only be negative, however, due to the fact that edges within digital aerial images are not well-defined entities.

Of course, if a mathematical model of an edge is assessed, then edge extraction performance can be analyzed. Virtually all literature reports of such analysis, however, concentrate only in modeling edge detection performance. This problem is analytically tractable since edge detection is a local process which is essentially non-interpretive. For example, Abdou and Pratt (Ref. 21) analyzed and compared the performance of popular 2×2 and 3×3 edge detection masks according to their ability to accurately measure the orientation and location of an edge. The model Pratt used was an ideal step edge under noise-free conditions. In an analysis of optimal digital filtering for edge detection (Ref. 22), Modestino and Fries modeled noise as a homogeneous zero-mean random process. Such approaches are necessary if one is to show how edges can be detected optimally or evaluated quantitatively. Unfortunately, there is no evidence that such models are appropriate for detecting edges in aerial imagery.

If modeling edges locally for the purpose of quantitative element detection analysis is somewhat artificial, formulating global models for quantitative evaluation of linking performance is even further removed from realistic conditions. Thus, such literature reports are relatively few, and those which have been reported (e.g., Ref. 23) have not been popular practical implementations.

Since edge extraction is not a technique which will permit the fully-automatic extraction of any arbitrary feature, comparative assessment of techniques with respect to feature extraction performance can only be accomplished within the framework of a semi-automated operational scenario. Since edges are not well-defined mathematical entities, evaluation of which techniques are best can only be accomplished by experimentation.

In conclusion, it is possible to say that for the purpose of extracting general types of edges, i.e., those which are generated by boundaries of the set of all types of features and therefore do not have predetermined shapes, an edge extractor which accomplishes the goals of each of the basic phases should be employed. Its linking phase should not presume some a priori line shape. An example of this set of general-purpose edge extractors is the *Navatia-Babu line finder (Ref. 19)*. Experimentation must be performed, however, to optimize the details for semi-automated feature extraction.

2.2.3 Segmentation

Segmentation techniques typically refer to region-finding algorithms, the class of techniques reviewed in this section. As mentioned in the overview, region-based segmentation (together with edge extraction) belongs to the class of low-level vision algorithms since its goal is to perform perceptual organization of an image. The goal of most segmentation algorithms is to completely partition the image into homogeneous regions. Thus, such algorithms are based on the assumption that images may be modeled as a set of such regions. Furthermore, there is an assumption that each region will correspond to some meaningful physical entity, like the projection of an object surface; otherwise, the segmentation would not be

useful. Because of these strong assumptions, such algorithms are interpretive. They force an interpretation of the image independently of whether the underlying image model has a well-defined relationship to physical object properties.

Homogeneous implies that each region must be perceptually uniform, i.e., that all parts of the region should appear to be essentially the same gray shade. However, in reality, real regions of significant size never have constant gray level. Thus, an algorithm must base region decisions on the average or statistical behavior of pixel intensities. The two primary classes of segmentation algorithms - region-growing and region-splitting - are discussed below. Survey papers on segmentation techniques include Refs. 3, 24 and 25. Segmentation of natural scenes is analyzed in detail in Ref. 26.

Region Growing - Region-growing algorithms typically begin by finding atomic regions of just a few pixels each. Each pixel within an atomic region has the same (or nearly the same) intensity. Thus, atomic regions are conservatively selected. Atomic regions are then grown by adding pixels which do not deviate significantly from the average intensity of the region. Regions may be grown until they reach each others borders, but this technique will not yield satisfactory results unless the regions are quite simple. Otherwise, decisions about where to place a pixel whose value is mid-way between the averages of adjacent regions are arbitrary. A more sophisticated technique (Ref. 27) is to grow regions until they share a common boundary, and attempt to merge two or more of them into a larger region. The merge decision may be based on region shape as well as average intensity. If the shape of the merged region will be simpler than the shapes of the two before merging, the merge is encouraged. Another technique (Ref. 28) uses semantic criteria to assist in making a merge decision. It assigns conditional

probabilities to regions (representing the likelihood that regions of the same interpretation will be adjacent) and makes merge decisions accordingly. Of course, for complex imagery, making proper interpretations and obtaining meaningful probabilities is extremely difficult.

Region Splitting - While region-growing follows a bottom-up approach to segmentation, region-splitting entails a top-down approach. Region-splitting begins by taking a histogram of a large image area. It then analyzes the histogram, looking for large peaks, which presumably indicate regions of pixels having approximately the same intensity value. Based on thresholds determined from the peak analysis, the algorithm then finds connected regions composed of pixels within the selected range. If some of these regions are large, it can repeat the process by treating them as new image areas and taking their histograms.

The main difficulty with this procedure is that there is no particular reason to expect that pixels whose values fall within a specified intensity range will all be members of the same connected region. Thus, techniques for cleaning up small island-like regions are required. One such approach is the "split-and-merge" technique of Ref. 29. After splitting, the algorithm looks for regions which can be merged, based on criteria such as shape, size, and adjacency. If regions are not sufficiently homogeneous, they qualify for additional splitting.

A number of algorithms base their measures of homogeneity on multiple radiometric measures, such as are available with color and Landsat imagery. The Ohlander (Ref. 30) algorithm is one of the most commonly used, but relies on color imagery. It bases decisions about region-splitting on the

analysis of multiple histograms of redundant color components (e.g., red, green, blue, intensity, hue, and saturation). It picks the best peak of all histograms as its basis for segmentation. Other techniques employing multiple radiometric measures are Refs. 31 and 32.

Assessment - The multi-radiometric techniques just described may be quickly assessed by observing they are not applicable since only black and white imagery is assumed to be available. With respect to the other techniques, the results of our assessment are similar to those which discussed edges and their relationship to object properties.

The goal of the techniques reviewed above is to find homogeneous regions. The problem is that the term homogeneous is not well-defined in itself, and furthermore has no well-defined relationship to meaningful object properties. Since most segmentation techniques partition an image, it means they will always find homogeneous regions, no matter what the contents of the image. Aerial images will contain arbitrarily-small regions, and statistical averages will lose their meanings for such small numbers of pixels. Many surfaces have discolorations or specularities which prohibit their characterization in terms of homogeneous measures. The conclusion is that since regions do not reliably correspond to feature-characteristic properties, segmentation algorithms are not useful for feature identification. Since regions do not correspond reliably to projected areas of features, they are not useful for feature delineation. Moreover, experimental evidence does not promise otherwise (Refs. 2 and 33).

2.2.4 Texture

Texture is the term applied to image regions which are not of constant gray shade, but seem perceptually homogeneous due to intensity variations which change with apparent regularity. Machine-perceptual textures are based on average properties of image intensity over a region. In this respect, they are like region-based segmentation techniques. However, by themselves they do not attempt to segment or organize an image. Textures alone are only representations of average pixel behavior.

Texture representations have two essential components: measurements and histogram-related representations. The simplest measurement is the intensity of a single pixel. Characterization of the behavior of individual intensity collections are called first-order texture measurements. The representation is always related to the intensity histogram. Statistical representations are mean, variance, and skewness, for example, but any measurement of histogram behavior could be used including the histogram itself.

Second-order measures describe joint average behavior of two pixels displaced from one another by some fixed amount. A co-occurrence matrix (Ref. 34) is a two-dimensional histogram of all possible co-occurring values of two pixels separated by a fixed displacement.

Rather than pixel intensities it is possible to use other measures, like edge elements. Both the magnitude and orientation of edge elements can be histogrammed. Laws (Refs. 1, 33) generalized this concept by convolving an image with a number of linear masks, and using the energies of each as the texture representation.

Texture measures and representations may be useful for discriminating among textures, as long as the image presented during measurement is all of the same texture. A texture technique would in general be confused if part of its measurements corresponded to one texture, and part corresponded to some other type of imagery. Because textures do not perform perceptual organization, they are not appropriate techniques for feature extraction.

2.2.5 Symbolic Techniques

Higher-level vision represents and matches information symbolically. The primary difficulty with high-level techniques is that they depend on unreliable low-level vision techniques, and therefore are not currently promising for application to feature extraction.

Two-dimensional shape representations can be used either for objects which are essentially two-dimensional or which always appear in images so that a specific view always takes place. Alternatively, two-dimensional representations may be intermediate to higher levels. Gross representations include simple global measures like area or area/perimeter. To the extent that a shape is well-described in the frequency domain by the curvature variations around its perimeter, Fourier descriptors make sense. The medial axis transform is a skeleton-like representation, every point of which is midway between the two closest points on the (connected) two-dimensional shape.

The most popular of the three-dimensional shape representations are generalized cyclinders. These shapes are defined by the volume swept out by a specified cross-sectional shape

which travels perpendicular to a pre-specified axis. At their most general, the axes may be arbitrary space-curves, and the cross-section may change shape as it travels. Gradient space is a dual space for alternatively representing various entities from regular three-space. It has the property that three-space phases map to gradient space points. Other properties make gradient space useful for reasoning about three-dimensional polyhedra, and so is a useful device for understanding blocks-world images.

Two- and three-dimensional model shape representations may be compared or matched to empirically-acquired representations for the purpose of object recognition. Template-matching of two- or three-dimensional shapes may be performed by correlating a stored shape with areas of imagery expected to contain that shape. Unless the shape is well-preserved in the imagery, however, template matching is bound to be unreliable. Another matching technique is graph matching. Shapes may be broken down into various components, each represented by a graph node, with interconnecting branches representative of component relationships. An empirical model is then matched wholly or partially against the paradigm.

2.2.6 Statistical Pattern Recognition

Statistical pattern recognition is based on statistical decision theory, and therefore on the notion that the characteristic elements of a pattern behave randomly on an individual basis, but can be characterized statistically as a group. A pattern (or object) is represented as a prototypical region in N-dimensional feature space. Each "feature" is a measurement of the image (e.g., edge-element magnitude). The idea is that, on the average, a set of measurements from a particular object will constitute a unique value, and thus correspond to a unique

region in N-space. If there are several such prototype regions, each corresponding to a different object, the N-space position of a sample measurement from an unknown object will be nearest to one prototype region, and the object thus recognized.

One major shortcoming of this approach is that it has no means of dealing with the perceptual organization problem. It was not stated above where the sample measurements are to come from. For black-and-white imagery, single pixel values do not make very useful measurements. Other measurements require greater image extent, and then there is no assurance that that extent spans only one object. The pattern recognition technique of unsupervised classification or learning is somewhat similar to perceptual organization, as it is not known in advanced how many classes there will be. The number of classes is determined by how many distinct clusters of measurements are found in N-space. In unsupervised learning problems, however, each set of measurements is known to have come from a distant object or pattern. Unsupervised learning is conceptually identical to the method of region-splitting according to clusters of multi-radiometric values in multi-dimensional histograms. It is different from other segmentation methods though, because multiple measurements are all from the same pixel, and thus are guaranteed to be a part of the same object or region. As long as black-and-white imagery is used, global image measurements are required, and statistical pattern recognition techniques are not applicable.

2.2.7 Syntactic Pattern Recognition

Syntactic pattern recognition focuses on the structural rather than statistical aspects of patterns. The structure may be the boundary of the object, or the repetitive pattern of object primitives (e.g., the texels used in structural analyses

of texture). The pattern classifier is a finite state machine (Parser) which processes input symbols (object primitives) and outputs terminal symbol(s) which signify the recognition of pre-determined object classes. In this sense, its motivation is similar to that of two-dimensional symbolic shape representations. However, it represents variations among pattern instances from the same class as different strings of pattern primitives which obey a set of grammatical rules. As with two-dimensional shape representations, it cannot currently be exploited for feature extraction because it relies upon low-level, perceptual organization processes.

2.3 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

2.3.1 Summary

This chapter has reviewed and assessed the current state-of-the-art in black-and-white feature extraction technology as it applies to the global DMA feature extraction process. Following a brief overview of candidate approaches for performing technique assessment, a particular approach - the IU paradigm - was described and its application to technique assessment discussed. The IU paradigm evaluation technique was then applied to several major classes of feature extraction techniques including edge extraction, segmentation, texture, statistical and syntactic pattern recognition and symbolic matching.

2.3.2 Conclusions

The approach selected for assessing candidate feature extraction techniques is to analyze the general machine visual-

perception problem in terms of the IU paradigm. This paradigm is a model for showing how the data within an image must be interpreted (in terms of its relationship to the physical properties of the objects it portrays) in order to permit the accurate inference of descriptive object properties. It also shows that the ability to solve complex recognition problems like *feature identification hinges on the ability to infer such* object properties reliably.

The IU paradigm focused on the fact that local surface information, the information that fundamentally characterizes object appearance, is lost in the image formation process. Thus, recovery of physical properties must therefore proceed on the basis of global image properties. However, it was shown that to conclusively infer an object property from a global image property requires severe image acquisition, object orientation and property constraints. *These constraints had parti*cularly strong impact upon region-based segmentation techniques. Since many of the latter techniques find regions which are characterized by some measure of signal uniformity, yet there is no clear relationship between such measures and actual object properties. However, while image regions appear to have no clear relationship to feature properties, image edges may be generated by physical phenomena of interest, including object boundaries. As a consequence, edge extraction techniques appear promising for the delineation problem. However, since image edges may not be well-defined due to the high level of detail of aerial imagery and its discrete nature, determination of the "best" edge extraction technique(s) to apply can only be accomplished within the scope of an operational scenario.

Also addressed were texture, pattern recognition and symbolic methods. Textures did not appear to be sufficiently robust due to their intrinsic statistical characterization of

regions which seldom are satisfied in reality. Pattern recognition techniques suffered from the same weakness. Symbolic techniques do appear promising, but are not currently useable because they rely heavily upon lower-level edge extraction and segmentation techniques.

Finally, it was observed that an important requirement for machine perception is perceptual organization, the problem of choosing which local portions of an image belong to the same object or feature. It was also observed that this problem is similar to the feature delineation problem, for which it is not necessary to identify the semantic class of a feature in order to find its boundary.

2.3.3 Directions for Further Research

The recurring theme throughout this chapter has been that low-level techniques are the weak link in the machine perception hierarchy. There has been recent promising work in several areas, however, which indicates progress towards overcoming those weaknesses.

The first area involves relating image information to the nature of object properties as they are instantiated via the image formation process. Witkin (Ref. 35) showed that the behavior of the autocorrelation function of a window as it slides across on edge boundary may indicate whether the boundary is an occlusion or a shadow. The autocorrelation function across a shadow boundary will generally be smooth, because only the mean intensity changes and is averaged out by the correlator. The partially-shadowed region is likely to be otherwise homogeneous.

A second example is an attempt to define a texture representation which is physically and mathematically justified. Pentland (Ref. 36) showed how a mathematical function for randomness is related to the non-deterministic behavior of the undersampling of objects in aerial images, which creates detail phenomena.

The second area focuses specifically on the problem of perceptual organization. Lowe and Binford (Ref. 37) suggest that recognition is likely to be intimately involved with low-level perceptual organization, and propose general principles that govern the grouping process. Fischler (Ref. 38) has developed an image line-finding algorithm which is claimed to be capable of finding image lines (not image contour locations which correspond to physical edge phenomena) as well as humans can. It should be emphasized that this comparison is best made when the lines appear to be without familiar context, so that human viewers do not make knowledge-based interpretations.

The research directions described above are consistent with the needs indicated in this chapter, that low-level techniques must address problems of perceptual organization, and that meaningful low-level techniques should relate image properties to the instantiated physical properties of objects. Further emphasis and support of these research directions would appear to be particularly beneficial.

3. CONCEPT OF OPERATION FOR A SEMI-AUTOMATED FEATURE EXTRACTION SYSTEM

In this chapter, a concept of operation for an interactive, semi-automated feature extraction system based on current technology is formulated. Figure 3-1 depicts TASC's initial concept for a semi-automated feature extraction system in the context of the overall DMA production process.

Source selection/digitization and image orientation are associated with the source package preparation, data base management, photogrammetric control generation, elevation data collection and control manuscript compilation functional areas of the DMA production centers. They are explicitly not assumed to be part of the feature extraction functional area. Image enhancement, area screening, and the detection, location, measurement, and classification of features are included in the feature extraction functional area, and are the subject of this concept of operation.

Since there are many possible implementations of the functions represented by the blocks in Fig. 3-1, several key factors had to be considered:

- Source selection/digitization and image orientation for feature extraction are relatively well-defined. These functions are already partially automated in such systems as the TA3/P, UNAMACE, ACE, AS-11, and RPIE. More advanced systems are currently under development (e.g., the Phase II data base system, the source assessment system, CAPI, DSCC)

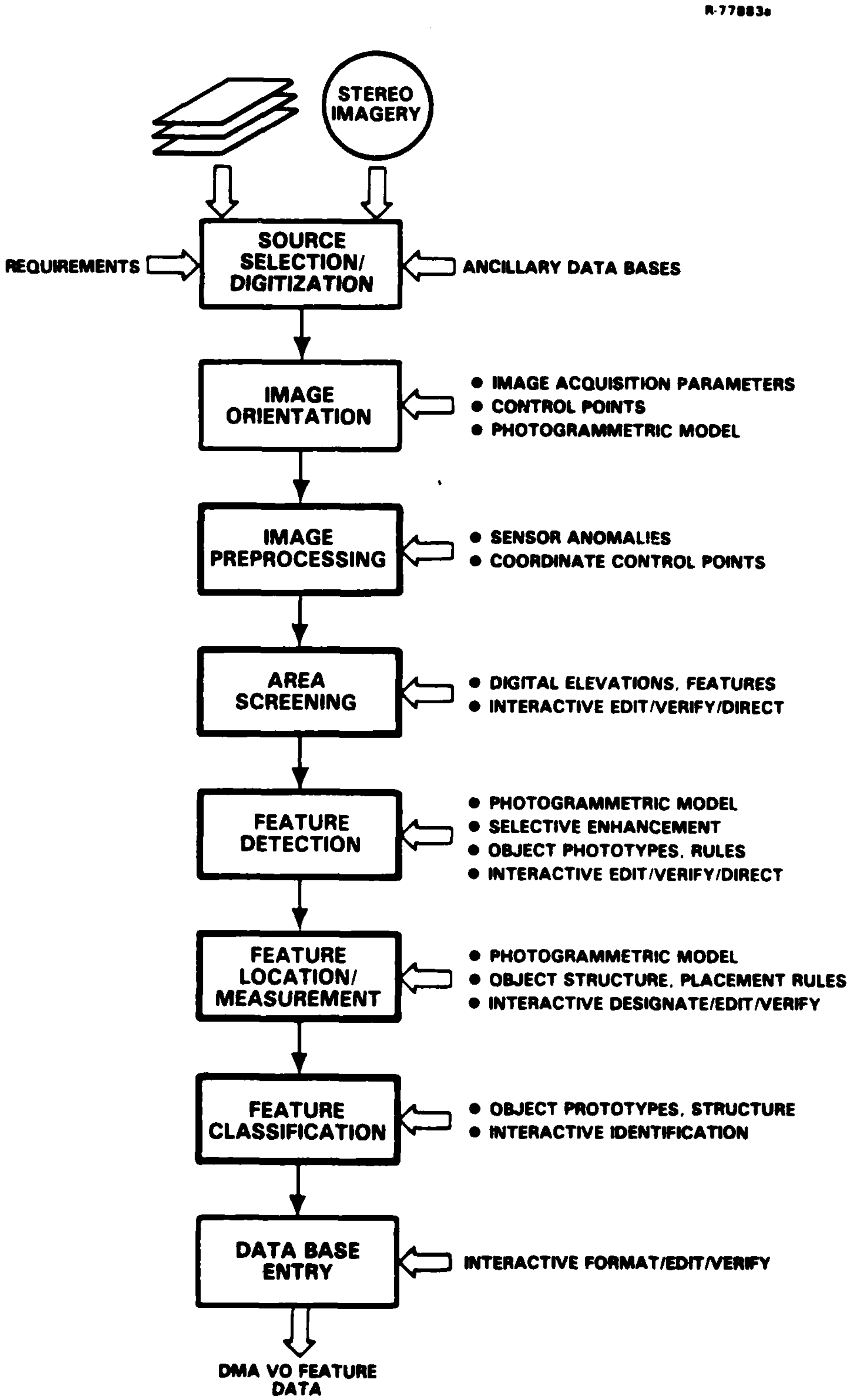


Figure 3-1 Initial Concept for Semi-Automated Feature Extraction

- Data base entry, apart from the larger data base management issues, is also a well-defined technology; advanced systems are contemplated (e.g., Phase II data base system, EDET adapted for feature editing) that will support these functions
- Area screening and feature detection, location, measurement, and classification are performed by visual scanning today. Visual scanning will continue as the primary mode of operation in such planned digital systems as CAPI and DSCC. Increased automation of this process could prove beneficial, but requires the cost-effective use of proven enhancement, pattern recognition, and knowledge-based feature extraction techniques
- Although candidate digital feature extraction techniques are available, they must be adapted and integrated into the DMA production environment.

In Task 3, numerous techniques were investigated in such areas as image processing, pattern recognition, and image understanding that could potentially contribute to automating the area screening and feature detection, location, measurement and classification functions. While some of these techniques were shown to be capable of automatically detecting or identifying features in highly constrained situations (e.g., small area, high-resolution, noise-free imagery), <u>no reliable, fully-automated detection techniques</u> suitable for performing general feature extraction were found. Furthermore, the nature of the feature extraction problem and of these technologies is such that <u>a fully-automated solution is not likely to be available in the near future</u>.

Thus, the best approach to developing the concept of operations was to assume a semi-automatic implementation which

made use of available feature extraction techniques to improve the baseline feature extraction process. Improvements would result mainly through the use of

- Interactive image enhancement to permit an image analyst to more easily see, measure and classify features of interest
- Semi-automatic screening to determine whether or not features of interest are likely to be present in selected areas
- Partial feature detection, which would cue operators to areas requiring further analysis, and cut down on the likelihood of undetected features
- Highly local, directed scene analysis coupled with photogrammetric models to locate, delineate and measure detected features
- Highly local, directed pattern recognition or feature desciption expert systems to classify features.

Note that each of these improvements corresponds to a function in Fig. 3-1. Realizing these improvements will require a careful selection of candidate techniques, and an objective evaluation of their performance in the context of a baseline manual process.

Recognizing the constraints of the problem, and the limitations of current feature extraction algorithms, computer technology, and DMA feature extraction operations, the concept of operations for a semi-automated feature extraction system was formulated as follows:

- A generic, implementation-independent concept of operation for feature extraction was developed to facilitate the

identification of functions and activities which could potentially benefit from automation in general and machine perception technology in particular.

- Techniques and technology assessed in Task 3 were related to selected activities within the generic operational concept, and alternative approaches for automating them were described

- A feasibility and cost/benefit analysis was performed for each proposed approach and the most promising methods for automating the selected feature extraction activities were identified

- A refined concept of operation was formulated based on the results of the latter analysis, and a candidate system architecture identifying possible hardware/software components for implementing the concept of operation was developed.

The remainder of this chapter is organized as follows. Section 3.1 describes a generic concept of operations for the feature extraction process. Section 3.2 then identifies feature extraction activities within the generic concept of operation which are candidates for automation and the application of machine perception. The potential application of machine perception to feature detection, identification and delineation is addressed and feasibility issues and cost/benefit trade-offs surrounding the use of machine perception techniques for feature extraction are discussed. Based on the results of Section 3.2, a refined concept of operation for a planimetric feature extraction system is presented in Section 3.3, detailing those areas which have been selected for the application of machine perception versus strictly manual assistance.

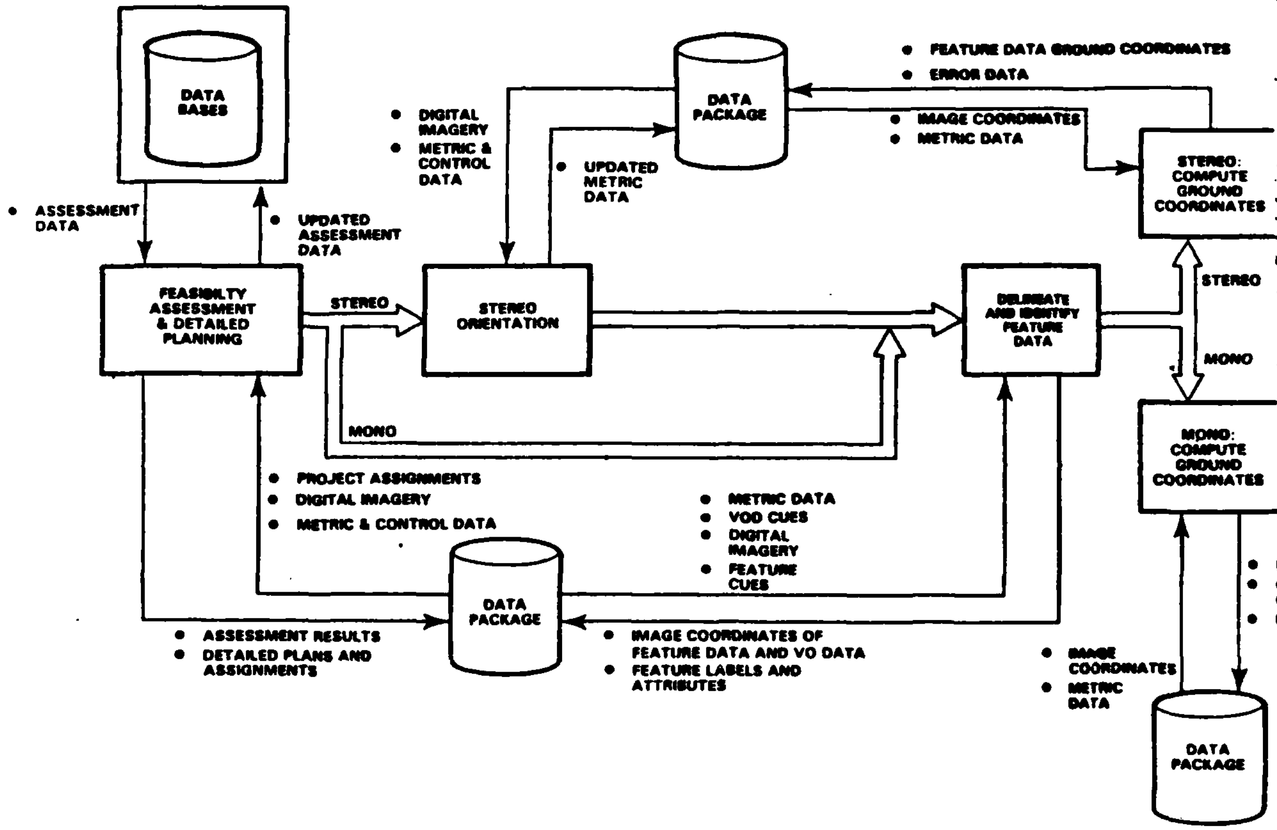
3.1 GENERIC CONCEPT OF OPERATION

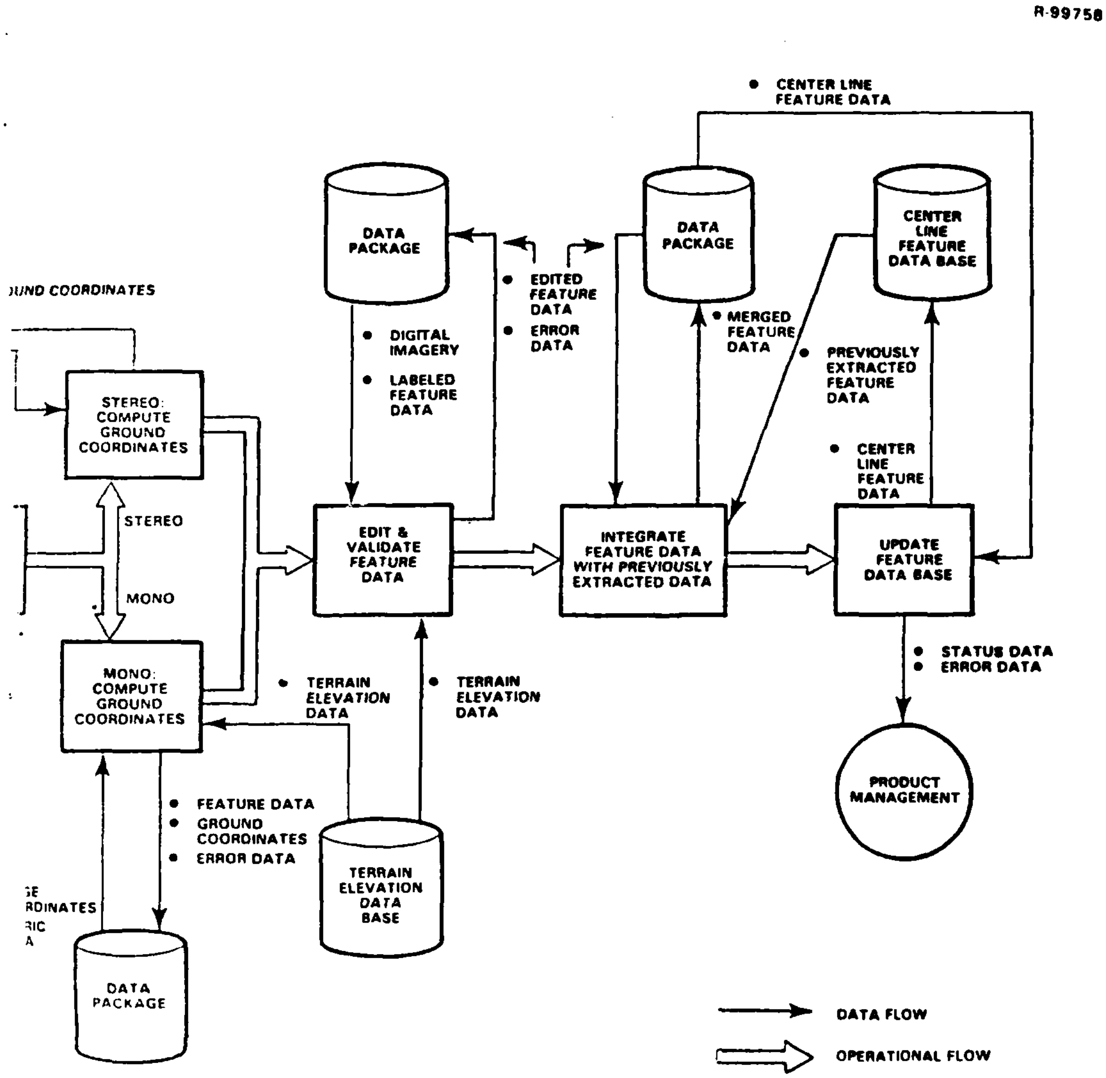
This section describes a generic concept of operation for feature extraction. The concept of operation describes the feature extraction process in terms of its component tasks and activities and is illustrated in Fig. 3.1-1.

As shown in the figure, feature data can be identified and delineated in either a monoscopic or stereoscopic mode. The computation of ground coordinates is different for each of these modes. For stereo delineation, ground coordinates are computed based on stereo intersection algorithms. For data which is delineated monoscopically, ground coordinates are computed using a single ray intersection with a ground model represented by a digital terrain elevation grid. Thus, the monoscopic mode is predicated on the existence of a terrain elevation matrix of sufficient density.

With techniques similar to those used for terrain elevation data extraction, the feature data may be overlayed on the operational image(s) and edited. The feature data is checked against existing terrain data for topographic and collection errors. These checks are accomplished both visually and through automatic statistical validation. As with terrain data, the feature data is integrated with previously extracted data to smooth the transition between models and geocells. The feature data base is then updated with formatted, merged center-line feature data. Detailed descriptions of these activities are provided below.

Figure 3.1-1 also shows a "data package" supporting feature extraction operations. The data package is a means of associating project data to meet the objective of rapid access for feature extraction operations. The data package is an


DATA BASES
ASSESSMENT DATA
UPDATED ASSESSMENT DATA
FEASIBILTY ASSESSMENT & DETAILED PLANNING
STEREO
MONO
STEREO ORIENTATION
DIGITAL IMAGERY
METRIC & CONTROL DATA
UPDATED METRIC DATA
DATA PACKAGE
FEATURE DATA GROUND COORDINATES
ERROR DATA
IMAGE COORDINATES
METRIC DATA
STEREO: COMPUTE GROUND COORDINATES
DELINEATE AND IDENTIFY FEATURE DATA
STEREO
MONO
MONO: COMPUTE GROUND COORDINATES
PROJECT ASSIGNMENTS
DIGITAL IMAGERY
METRIC & CONTROL DATA
METRIC DATA
VOD CUES
DIGITAL IMAGERY
FEATURE CUES
DATA PACKAGE
ASSESSMENT RESULTS
DETAILED PLANS AND ASSIGNMENTS
IMAGE COORDINATES OF FEATURE DATA AND VO DATA
FEATURE LABELS AND ATTRIBUTES
IMAGE COORDINATES
METRIC DATA
DATA PACKAGE



2

Figure 3.3-1 Generic Feature Extraction Concept of Operation

assemblage of all data and/or references to data that are required for a given project. A key feature of the data package concept is that the data package may contain either the actual data or indices that point to the data in the data bases. The function of the data package is to provide data where and when it is needed and in the required format.

Eight major activities comprise the feature extraction process as illustrated in the figure:

- Feasibility Assessment
- Stereo Orientation
- Feature Identification and Delineation
- Stereoscopic Computation of Ground Coordinates
- Monoscopic Computation of Ground Coordinates
- Editing and Validation of Feature Data
- Integration of Feature Data with Previously Extracted Data
- Feature Data Base Updating

The feasibility assessment activity consists of receiving project assignments from production management (including products/areas of interest, project priorities and project schedules), and developing detailed production plans for the feature extraction process.

If stereo imagery is used for compilation, stereo orientation will be required which includes assessing the data package in order to extract stereo imagery and its metric control data, and subsequently establishing the stereo model. If monoscopic imagery is used, a similar set-up procedure is required.

The key feature extraction activity consists of identifying, delineating and labelling feature data (in either mono or stereo). Rectification (to assist in stereo fusion of imagery), imagery enhancement and interactive machine perception techniques may be used to assist in identification, delineation and labelling as appropriate.

Since one of the primary outputs of feature extraction is the geographic coordinates associated with each feature, computation of such coordinates is required. In the case of stereo feature extraction, geographic coordinates may be computed directly using stereo intersection techniques. In the case of mono feature extraction, geographic coordinates may be obtained by accessing available terrain elevation data and computing the coordinates by interpolating within the terrain elevation data matrix. Depending upon feature location accuracy requirements, the latter method may or may not yield sufficient accuracy.

When feature extraction has been completed, the data must be edited to eliminate discrepancies and inconsistencies among contiguous features. The editing process is intended to filter mismatches, duplication and inconsistencies in the labelling, placement and location of features.

The final two steps of the feature extraction process are the integration of newly extracted feature data with previously extracted data, and updating the feature data base.

3.2 APPLICATION OF MACHINE PERCEPTION TO FEATURE EXTRACTION

This section discusses those feature extraction activities described within the generic concept of operation which are candidates for automation and the application of machine perception. Although the basic purpose of this effort is to identify and apply all forms of technology that could contribute to automating the feature extraction process (and thereby increase its productivity), this section focuses on how best to automate the processes which currently require human visual perception; namely, feature identification and delineation. These are the processes which have resisted automation to date.

Feature extraction activities may be partitioned into four major categories:

- Visual perception activities
- Perceptual support activities
- Record keeping activities
- Pre- and post-processing activities.

Perceptual activities are the activities which directly require a visual perception capability. They include specifically, feature identification, delineation, and feature Descriptor recovery. Any automation of these activities will be termed machine perception. The prospect of achieving full automation of these activities using current pattern recognition/computer vision technology was the topic of Task 3.

Perceptual activities break down into the two subcategories of identification (or recognition) and delineation

(or mensuration). Identification includes both the process of recognizing features according to their semantic category as defined by the list of FID codes, and the processes for recovering such semantic type descriptors as surface material category, roof descriptor, shape code, and the various micro-descriptors.

Delineation or mensuration activities are so designated because they all require some form of measurement (at least in a broad sense). A more accurate description of these activities might be perceptually-guided demarcation activities. Mensuration activities are directly involved with the recovery of numerically-valued Descriptors as well as feature delineation activities. All of these activities require the recovery of measureable physical properties of features.

Perceptual support activities directly assist the human analyst in the performance of the perceptual activities, but do not themselves execute the perceptual tasks. Perceptual support activities can be subdivided into machine assistance for identification (of feature and other semantic Descriptors) or for delineation (of feature boundaries or the measurement of numerical Descriptors).

There are not many ways in which an analyst can be assisted in the perceptual process of making identifications. Machine assistance for identification will be defined to include any automated process which directly enables the analyst to better identify features. Indirect activities which take place before the perceptual task begins (e.g., sensor data preprocessing and source selection) are excluded. Methods which do qualify are exemplified by stereo image display, magnification (or "zoom"), interactive image-enhancement techniques, and expert/advisory feature description assistance.

Since delineation is essentially mensuration, which requires hand-eye coordination to mark an image while visually picking-out the point of interest, all feature extraction systems will provide some kind of delineation assistance. On a stereocompiler, for example, machine delineation assistance refers to the system of controls and positioning mechanisms that move and measure images relative to one another to bring a conjugate coordinate-pair into correspondence. In a digital workstation, it is the system and software which include a user input device, graphical cursor and delineated-contour overlays, and graphics-editing techniques.

Record keeping activities are those activities involved with recording extracted DFAD information. DFAD includes all Descriptors within the Feature Header Record and the set of feature boundary coordinates. All Descriptor entries are recorded as numbers, but as discussed above, are numerical in nature. Those which are semantic-valued are converted to a pre-defined code. For each possible semantic value, there is a unique code specified by the DLMS document. Thus, two record-keeping tasks become apparent: the process of recording the DFAD information once extracted, and the process of translating semantic values to their numerical codes. Conceptually, the translation process is simply one of table lookup. If identification were being performed automatically by machine perception, the semantic value of a Descriptor would be represented within the machine as a unique symbol, and in as much as all represented machine symbols can be interpreted as numbers, the symbolic representation would be equivalent to the unique Descriptor code. Thus, when automated identification becomes feasible, translation to numeric code will be implicit. For the experienced human analyst, translation will also become an implicit process, at least for Descriptors with a small number of possible values. For FID Descriptors, however, there

is a large and potentially ever-growing list of features, and even an experienced analyst will occasionally have to refer to a translation rule to obtain the code for a new or rarely-encountered feature.

All of the tasks related to record keeping are highly amenable to automation, and can be conveniently accessed by an analyst from an on-line interactive data base system as a component part of the workstation. Such capabilities have been defined and implemented at the prototype level by the IFASS system at DMA. As such, they will not be further treated in this report.

Finally, there are various pre- and post-processing activities which are vital for achieving the end result of obtaining DFAD for all features within a manuscript region, but which are not closely enough related to the perceptual activities to be studied in Tasks 3 and 4. These include such preprocessing activities as establishing controlled imagery (i.e., obtaining the parameters that relate geographic and image coordinate systems through the camera model), and computational coordinate-conversion processes (e.g., finding the earth coordinate for a conjugate image coordinate pair, and converting to image coordinates from a mensuration instrument's coordinate system). Also included are such post-processing activities as the digitization and editing of feature boundaries (for the manual delineation FE process), and merging of extracted manuscript data in a consistent fashion with spatially-adjacent or pre-existing manuscripts.

In the remainder of this section, possible methods of automating the visual perception and perceptual support activities are described. Those classes of machine perception techniques which show promise for application to the feature

extraction process are identified in Section 3.2.1, and alternative approaches for applying the latter techniques to feature identification and delineation are addressed. In Section 3.2.2, the feasibility issues and cost/benefit trade-offs associated with the alternative approaches proposed are discussed.

3.2.1 Approach to Machine Perception for Feature Identification and Delineation

This section discusses possible approaches to automating the feature extraction process, in particular, on the application of machine perception technology to those activities comprising feature identification and delineation (Fig. 3.1-1). According to the CAP Standards (Ref. 108), these activities comprise a significant portion of the time currently spent on feature extraction. Moreover, unlike many of the other functions and activities associated with feature extraction, these activities have most strongly resisted automation to date.

For the purpose of this discussion, we assume that the feature identification and delineation processes consist of the five activities shown in Fig. 3.2-1.

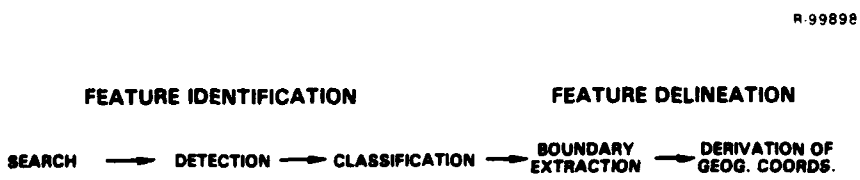


Figure 3.2-1 Feature Identification and Delineation Processes

The first three activities - search, detection and classification - are typically associated with feature identification, while the latter two activities are associated with feature delineation. Note that although the activities appear to be

sequential as portrayed in the figure, there is no reason for them to be so in an actual operational environment. In the sections below, alternative ways in which each of these activities may be automated are discussed.

Semi-Automation of Feature Identification - As mentioned above, feature identification consists of three fundamental activities. The first is search wherein an imagery analyst (IA), working with either monoscopic or stereoscopic imagery, systematically scans his imagery to ensure that he detects all features of interest to the particular product for which he is responsible. Since the IA does not typically know beforehand whether or not a region of imagery contains any features of interest, he must always assume the worst (i.e., that all regions contain features of interest) and consequently search all regions of the imagery. Depending upon the complexity of the scene the IA is viewing, it is possible that as much time will be spent searching regions that contain no features of interest as will be spent searching regions that do contain such features. This poses a twofold problem since not only is the IA's effective throughput reduced because he is spending time over regions that contain no information of interest, but his fatigue factor is also increasing, which may reduce his throughput (and accuracy) searching regions that do contain information of interest. Thus, one goal of semi-automating the feature extraction process should be to reduce the amount of time spent in searching regions that contain few if any features of interest.

The second fundamental activity in feature identification is feature detection; i.e., recognizing a feature occurring in a region of imagery currently being searched as a feature of interest. The difficulty of the recognition process is proportional to the complexity of the scene being viewed,

the complexity of the feature of interest, and the resolution and noise-freeness of the imagery. Despite these variables, preliminary recognition* of a feature can generally be accomplished using only a few feature attributes or cues to select/distinguish it from other features. As a consequence, even in human beings, this phase of recognition can lead to false detections which must subsequently be eliminated by more focused, higher level recognition processes. The mechanisms by which an IA (and humans in general) performs higher level recognition are themselves not well understood and it does not appear likely (given the results of the survey of feature extraction techniques conducted in Task 3) that this phase of recognition will be automated in the near future. However, it does appear that some form of automation of the preliminary recognition phase may be possible. Thus, another goal of semi-automating the feature extraction process should be to support the process of preliminary recognition on behalf of feature detection.

The third fundamental process in feature identification is that of feature classification; i.e., obtaining the feature name and associated descriptors for each feature of interest that has been detected. Classification is performed in accordance with the DLMS V Specification (Ref. 4) and is complicated by such factors as imagery resolution and quality, the IA's knowledge of and familiarity with both the Specification and the various possible manifestations (instantiations) of each feature of interest within an image, fatigue, and scene complexity. Since it has been shown that it is highly unlikely in the near future that any fully-automated support for the classification process itself will be available, another goal

*Preliminary recognition is defined as the state of initial awareness of a feature in an image as a potential feature of interest.

of semi-automating the feature extraction process should be to provide (as required) support for the feature classification process by providing expert feature description assistance, and reducing the possibility of errors introduced by the complicating factors mentioned above.

Semi-Automation of Search and Preliminary Recognition - As stated earlier, the goals of a semi-automated capability for search and preliminary recognition are to reduce (or eliminate) the amount of time spent in searching regions that contain few, if any, features of interest, and to assist the process of preliminary recognition in order to support overall feature detection. These two goals are tightly-coupled, and assume that there exists, or can be defined, a small number of attributes or cues for each feature which are:

- invariant for all image acquisition and scene conditions
- always visible or inferrable when the feature of interest is present
- derivable from the signal properties of the image alone
- sufficient to permit preliminary recognition of a feature or features of interest*.

In the following, such attributes will be referred to as the minimal attribute subset (MAS) of a feature. The definition of this minimal attribute subset is clearly feature-dependent, and has not been identified to our knowledge for any major class of DLMS features to date. However, should such a subset be identified, it could be used as shown in the scenario portrayed in Fig. 3.2-2 to facilitate the search and preliminary recognition processes.

*But not necessarily sufficient for absolute recognition.

R-99899

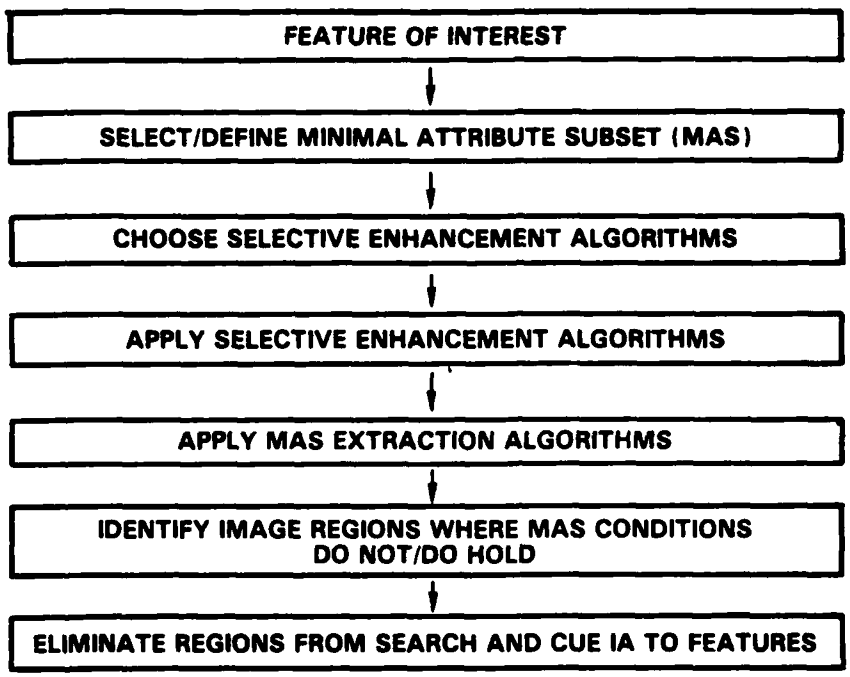


Figure 3.2-2 Semi-Automated Search/Preliminary Recognition Scenario

The scenario begins with an IA informing the feature extraction system about those features he is interested in extracting. For each such feature or class of features, the system would then retrieve the associated minimal attribute subset, and use it to invoke and parametrize, either in parallel or sequentially, image processing algorithms specifically designed to enhance those attributes contained within the imagery. After the latter selective enhancements had been accomplished, a set of minimal attribute detection algorithms could be run over the image (or images, since a number of intermediate forms of imagery might be generated) to determine the presence or absence of the MAS attributes over pre-defined regions within the image. Subsequently, a simple production

rule or expert system could, for each region of the image, determine whether or not the appropriate conditions have been satisfied for the presence of the MAS to be established. If they do not, the system would designate (e.g., graphically) those regions that do not or are highly unlikely to contain the feature(s) of interest. If MAS conditions do hold, the system would designate those regions that are likely to contain features of interest and graphically cue the IA to their possible locations.

Semi-Automation of Classification - Although the full automation of the classification process is highly unlikely in the near future, machine assistance to eliminate or overcome some of the difficulties associated with classification is possible now. The basic classification process involves assigning a name and a number of descriptors to each feature of interest detected. The specifications/definitions of these names and descriptors are numerous and sometimes complicated. Depending upon IA experience, the number of features present, image quality, fatigue and other factors, classification can become quite difficult. Although no automated system yet exists which will automatically classify features, capabilities in the form of expert or advisory systems resident within relatively low cost work station technology do exist and could be used to assist the IA in classifying features in a highly interactive fashion. Figure 3.2-3 illustrates a possible interaction. The dialog shown would continue until one or more hypotheses are eliminated, a single hypothesis is confirmed, the IA needs no further assistance, or the system can provide no further information.

The scenario could just as easily have begun with the user providing the system a description of the feature properties visible, and letting the system do the work of attempting

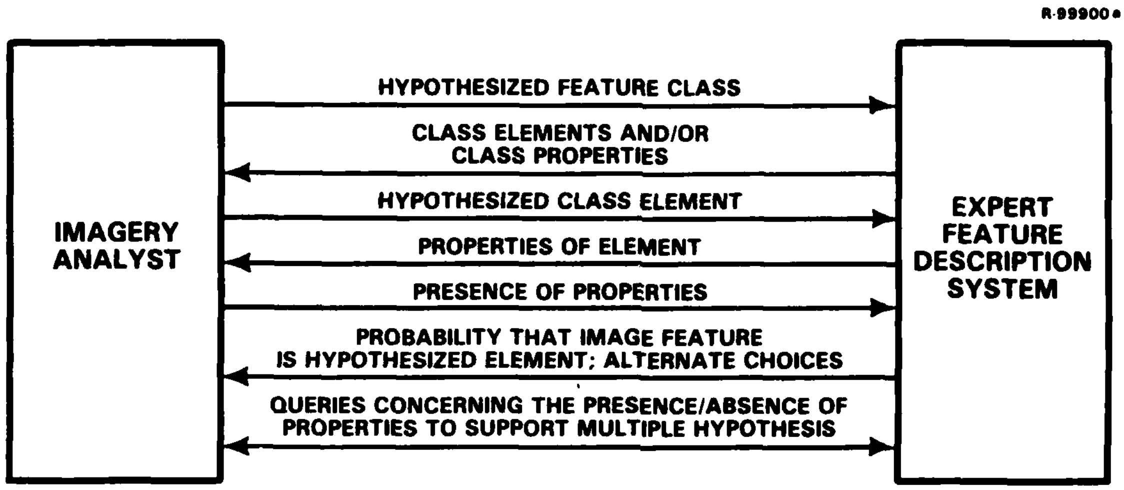


Figure 3.2-3 IA/Expert Classification System Interactive Scenario

to make the initial feature class and element hypotheses. In a digital work station environment the entire scenario of Fig. 3.2-3 could be supported on a single high resolution monitor using a multi-windowing presentation strategy with the only operator interaction with the system via a three-button mouse.

Semi-Automation of Feature Delineation - Feature delineation involves the identification and delineation of feature boundary points and the derivation of the geographic coordinates (latitude, longitude and height) associated with each point. Since delineation occurs after identification (at least perceptually), it is possible to bring more information to bear to assist an automated delineation process. Several semi-automated scenarios are possible for accomplishing (the boundary extraction portion of) delineation:

- Mono Boundary Extraction Scenario 1 - The IA identifies to the system those features he wants to delineate and the system, based on either the MAS for each feature or extensions to it, attempts to extract the boundaries of the features (using selective enhancement and extraction procedures) and graphically portrays the boundaries it has extracted as (registered) overlays on the operational image. Delineation is performed on a global image basis, and the results are presented to the IA for review and editing after the process is complete.

- Mono Boundary Extraction Scenario 2 - The IA identifies to the system those features he wants to delineate and approximately designates (by electronic grease pencil, polygonal region-of-interest, or other means) those portions of the operational image containing the features of interest. The system then applies selective boundary extraction algorithms to delineate the features contained within each designated region and displays the results graphically to the IA as an overlay on the operational image for subsequent verification and editing.

Clearly, the most computationally efficient scenario is that provided by scenario 2, since the boundary extraction computations are directed/restricted to image regions known to contain the feature of interest. Moreover, the processes of automatic boundary extraction and manual editing may take place in parallel, since regions whose boundaries have been extracted are immediately available for editing. The advantage of automation in this case is its ability to free the IA from having to perform precise delineation, and instead letting the IA perform rough delineation and final editing, both of which can be performed by the IA relatively quickly and require less demanding perceptual/motor skills.

The process of obtaining the geographic coordinates associated with the boundaries extracted above is less straightforward. Currently, two techniques predominate:

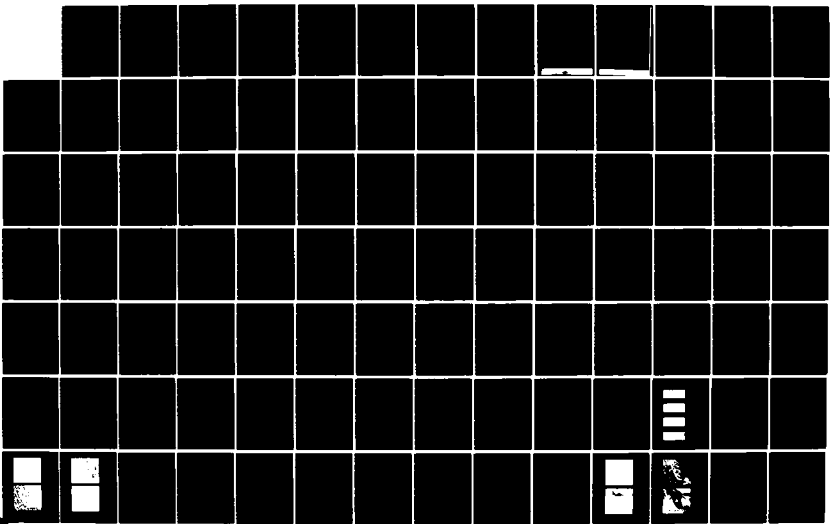
- Transfer of monoscopically extracted feature boundaries to orthophotos for subsequent digitization and control
- Delineation of boundaries in stereo so that geographic coordinate data may be extracted concurrently with the feature boundary.

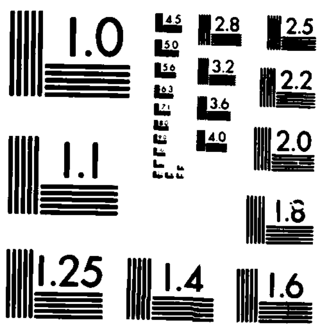
Another technique that could also be used would employ DTED in conjunction with monoscopic boundary delineation to obtain geographic coordinates via interpolation. Recalling our initial assumption, however, that no collateral information was assumed to be available (in particular DTED, but including orthophotos as well), the only viable scenario appears to be that of performing delineation in stereo. Two stereo mensuration scenarios present themselves:

- Stereo Mensuration Scenario 1 - The Monoscopic boundary extraction scenario 2 above is applied to both halves of a stereo pair. Following IA editing, the system then attempts to find corresponding points (e.g., by template matching along epipolar lines). A batch-like process is then employed to produce geographic coordinates.
- Stereo Mensuration Scenario 2 - Monoscopic boundary extraction scenario 2 above is applied only to one half of a stereo pair. Little or no editing takes place; instead, the boundaries delineated in the one conjugate image are used to drive a correlation process which attempts to identify the corresponding boundaries in the other image. (Note that the geometric models for each image will be required.) The results of the correlation will then be displayed graphically

AD-A150 189 FEATURE EXTRACTION ASSESSMENT STUDY(U) ANALYTIC SCIENCES CORP READING MA M J CARLOTTO ET AL. NOV 84 ETL-0377 DACA76-82-C-0004

UNCLASSIFIED F/G 8/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

> in stereo (a la CAPI) for subsequent editing. Following editing, a batch-like process will determine the geographic coordinates of the conjugate extracted boundaries.

Clearly, from a machine computational point of view, scenario 2 is more efficient. Moreover, since editing takes place in stereo, the accuracy of correlation and delineation is more easily determined. Therefore, scenario 2 is recommended above scenario 1.

3.2.2 Feasibility Issues and Trade-Offs

This section discusses a number of feasibility issues and cost/benefit trade-offs associated with the approaches to semi-automating the feature identification and delineation processes discussed in the last section. The discussion of feasibility issues centers about whether the assumptions and technology underlying the scenarios appear to be realistic and realizable in the timeframe under consideration. The discussion of cost/benefit trade-offs focuses on whether or not, given that a particular approach/scenario is feasible, it is also cost-effective with respect to the largely manual processes used now.

Feasibility Issues - This section discusses several assumptions and technology issues which weigh heavily in determining the feasibility of the semi-automated approaches discussed in Section 3.2.1. The assumptions and technology issues concern:

- Source Data
- Digital Workstation Environment

- Feature Extraction Techniques/Machine Perception Techology.

With respect to source data, the key issue is whether or not image quality will be sufficiently high to facilitate both an IA's interpretation of a scene and automated perceptual processes designed to support the IA's interpretation in the context of the scenarios discussed in Section 3.2.1. The specific elements influencing image quality include:

- Resolution (i.e., how much ground resolved distance a pixel in image space represents), which affects the level of feature detail discernable
- Scale variation (i.e., how much the ground resolved distance of a pixel changes over an image), which affects both how much feature detail is discernable, and the way in which image processing/registration algorithms are applied
- Noise (i.e., how much of the information in an image is due not just to the scene being observed but to sensor or other anomalies), which affects overall interpretation and can significantly affect the performance of feature extraction algorithms
- Solar elevation and azimuth, which impact the way in which a scene is illuminated relative to the collection system
- Image acquisition process (i.e., whether the image is acquired digitally or in analog form), which affects the dynamic range of the data to be processed.

The results of Task 1 and subsequent information gathered during the course of the study appear to indicate that the primary sensor envisioned for collecting imagery during

the timeframe will meet most, if not all, of the feature extraction processes requirements for image quality in the FY85 (and beyond) timeframe. Therefore, source data will not pose a feasibility problem.

With respect to the digital workstation environment, the key feasibility issues surround whether or not sufficient processing, storage, communication and display resources are or will shortly be available to support the highly computational, yet highly interactive feature extraction scenarios discussed earlier. Specific issues include:

- Storage - is technology sufficiently well advanced to support the storage of several operational images together with all extracted symbology and graphics in a workstation environment?
- Processing - does the technology exist to support all of the image processing, classification, feature extraction and expert/advisory system processing required?
- Communication - is communication technology in general, and local area networking technology in particular, sufficiently well advanced to support the image, graphics and collateral information transfer requirements of one or more feature extraction workstations?
- Displays - does the display technology exist to support the high resolution (i.e., $\geq$ 1K × 1Kpixel) display and manipulation (e.g., zoom, slew, rotate) of monoscopic or stereoscopic black-and-white images and color graphics?

The results of Task 3 indicate that, although some of the feature extraction workstation requirements exceed the current state-of-the-art, most can now be satisfied by commercially available, off-the-shelf hardware and software. For

example, a number of vendors (e.g., IBM, DEC, STC) now offer standard magnetic disk products with capacities in excess of 1/2 to 1 Gigabyte. Advances in mini and microcomputers (e.g., VAX-11/780, 11/785, MC68010 and MC68020) and special array processors (e.g., FPS 5000 series, CSPI MAP, and Star Technologies ST-100), which provide instruction speeds of from 1-10 million instructions per second (MIPS) and 10-100 MIPS, respectively, indicate that highly powerful processors can reside in an interactive, local work station environment. Moreover, a number of 10 Mbit/sec and greater commercially-available local area networks now exist (e.g., Interlan 10 Mbit/sec Ethernet, Proteon Assoc. 10 Mbit/sec token ring bus, Network Systems Corp. 50 Mbit/sec Hyperchannel) which could potentially more than satisfy the communication requirements mentioned above. Finally, high resolution displays providing in excess of 1280×1024 pixel resolution with 1-24 bit/pixel dynamic ranges are offered by numerous vendors ranging from Ramtek, Mitsubishi, Tektronix, and Conrac for basic monitor configurations to Symbolics, DeAnza, and Ikonas for sophisticated graphics and image processing applications. Our conclusion is, therefore, that the feature extraction digital workstation requirment present no feasibility problems.

With respect to feature extraction techniques and machine perception technology, the key feasibility issue centers about the lack of low-level or primitive feature extraction/vision operators that are both robust and accurate. As discussed in Chapter 2 , no operation currently exists that can consistently extract such feature primitives as regions and edges in a reliable and perceptually consistent and meaningful fashion (much less organize them into higher level constructs). Realizing this, the notion of a minimal attribute subset was introduced, which, although incapable of supporting absolute feature recognition, could distinguish in an approximate fashion

between different classes of features in a reliable and predictable manner. Whether or not such a subset can be defined for DLMS Level V features is still an open issue, and represents a key risk to the feasibility of the scenarios discussed in Section 3.2.1.

Cost/Benefit Trade-Offs - This section identifies several major trade-offs which weigh heavily in determining the cost/benefit of the semi-automated approaches discussed in Section 3.2.1. The trade-offs include:

- Machine time versus manhours - will the (semi-) automation of a particular feature extraction task improve its response time and lessen/eliminate the number of manhours required?
- Man/machine interaction efficiency - for those tasks where semi-automation appears feasible, will the expected performance improvments to be gained by allocating machine and human resources to complementary portions of the task be outweighed by performance degradation introduced by inefficient interfaces at those instants when man and machine must interact?
- Equipment versus labor costs - despite the fact that particular portions of the feature extraction process may be automatable and that overall feature extraction system performance may be improved, does the cost associated with providing this capability outweigh its performance benefits?
- Throughput Improvement - does the semi-automation proposed yield higher feature extraction throughput?
- Accuracy Improvement - does the semi-automation proposed yield higher quality and more accurate feature extraction so that the time required to review and edit the output is reduced?

All of these issues/questions must be answered in the context of a proposed system implementation and in conjunction with experiments which can quantify the relative utility of machine perception versus human interpretation in performing selected feature extraction tasks. However, a preliminary subjective comparison of a semi-automated feature extraction process (according to the scenarios outlined in Section 3.2.1) with a highly manual process is illustrated in Table 3.2-1. The results indicate that semi-automation is both beneficial and could be cost-effective as well.

TABLE 3.2-1
COMPARATIVE ASSESSMENT OF MANUAL vs SEMI-AUTOMATED FEATURE EXTRACTION

COST/BENEFIT TRADE-OFF	MANUAL FEATURE EXTRACTION	SEMI-AUTOMATED FEATURE EXTRACTION
Machine Time vs Man Hours (to perform same task)	Machine time = low Man hours = high	Machine time = low-moderate Man hours = moderate
Man/Machine Interaction Efficiency	low	high
Equipment vs Labor Costs	low	moderate
Throughput	low	moderate
Accuracy	low	moderate

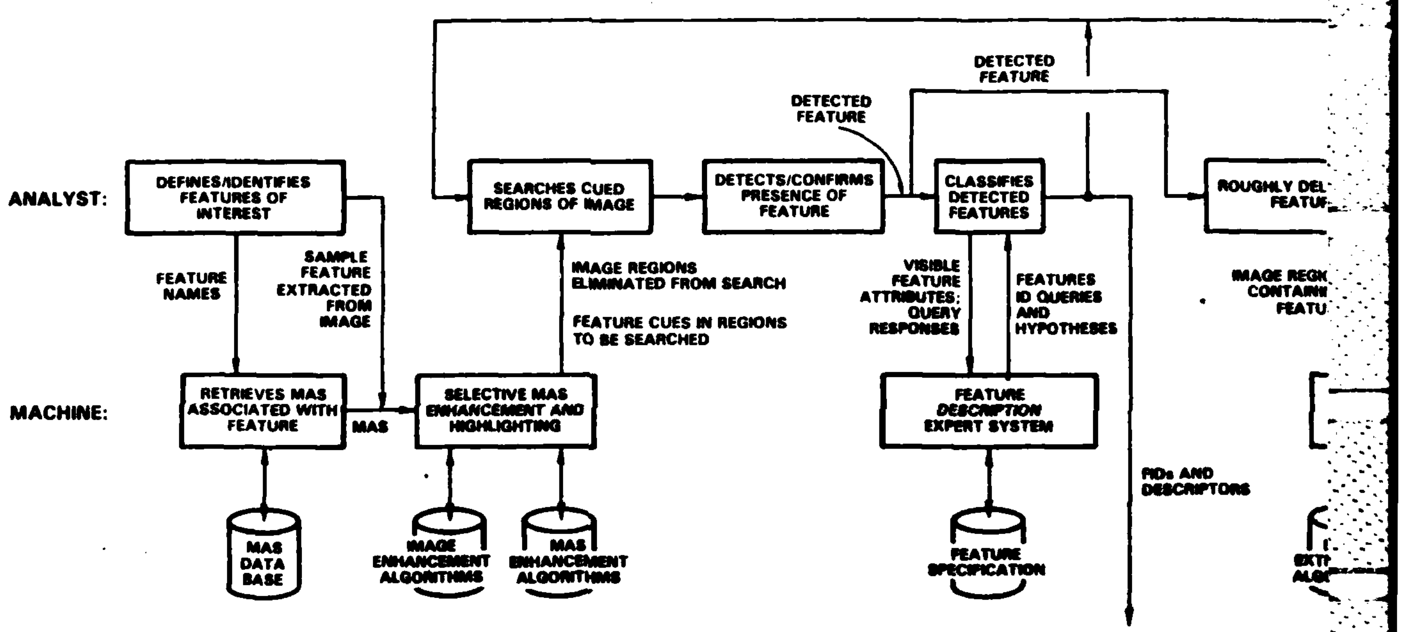
3.3 REFINED CONCEPT OF OPERATION

This section provides a refined concept of operation for a semi-automated feature extraction system which consolidates the more promising scenarios described in Section 3.2 for feature identific ion and delineation. A refined concept of operation for a semi-automated feature extraction system is

illustrated in Fig. 3.3-1, and parallels the generic description of the feature identification and delineation processes illustrated in Fig. 3.1-1. To briefly review the concept, an analyst first identifies to the system, either symbolically (via feature name) or through samples extracted from the imagery, those features he is interested in extracting. If a feature name is provided, the system assumes that it contains a minimal attribute set (MAS) for the feature, retrieves the MAS, and passes it to a module which selects image processing algorithms designed to enhance and extract the attributes within the MAS. If no MAS exists for a particular feature, then an actual sample of the feature as it appears in the image would be sent to the module. In either case, the system would then attempt to enhance/extract all occurrences of the MAS or feature sample in the operational image. Those regions of the images for which no MAS or feature were detected would be identified to the analyst via graphic overlay. Those regions for which a MAS or feature was detected would also be identified, and furthermore, all detected occurrences of the MAS or feature would be highlighted.

Based on this information, the analyst would begin a directed search through those image regions containing features of interest, and upon detecting a feature, would begin to classify and/or delineate it. With respect to classification, if the analyst was able to directly identify the feature ID (FID) and associated descriptors, he could input them directly into the system; however, should he require help in determining the FID code or any descriptors, a resident expert/advisory feature description system would be available to assist him.

Since feature IDs and descriptors are location tagged, a second and integral part of the process is delineating the boundary and determining the location of each feature. To aid



ANALYST:
DEFINES/IDENTIFIES FEATURES OF INTEREST
SEARCHES CUED REGIONS OF IMAGE
DETECTS/CONFIRMS PRESENCE OF FEATURE
DETECTED FEATURE
DETECTED FEATURE
CLASSIFIES DETECTED FEATURES
ROUGHLY DEL
FEATURE NAMES
SAMPLE FEATURE EXTRACTED FROM IMAGE
IMAGE REGIONS ELIMINATED FROM SEARCH
FEATURE CUES IN REGIONS TO BE SEARCHED
VISIBLE FEATURE ATTRIBUTES; QUERY RESPONSES
FEATURES ID QUERIES AND HYPOTHESES
MACHINE:
RETRIEVES MAS ASSOCIATED WITH FEATURE
MAS
SELECTIVE MAS ENHANCEMENT AND HIGHLIGHTING
FEATURE DESCRIPTION EXPERT SYSTEM
PIDs AND DESCRIPTORS
MAS DATA BASE
IMAGE ENHANCEMENT ALGORITHMS
MAS ENHANCEMENT ALGORITHMS
FEATURE SPECIFICATION

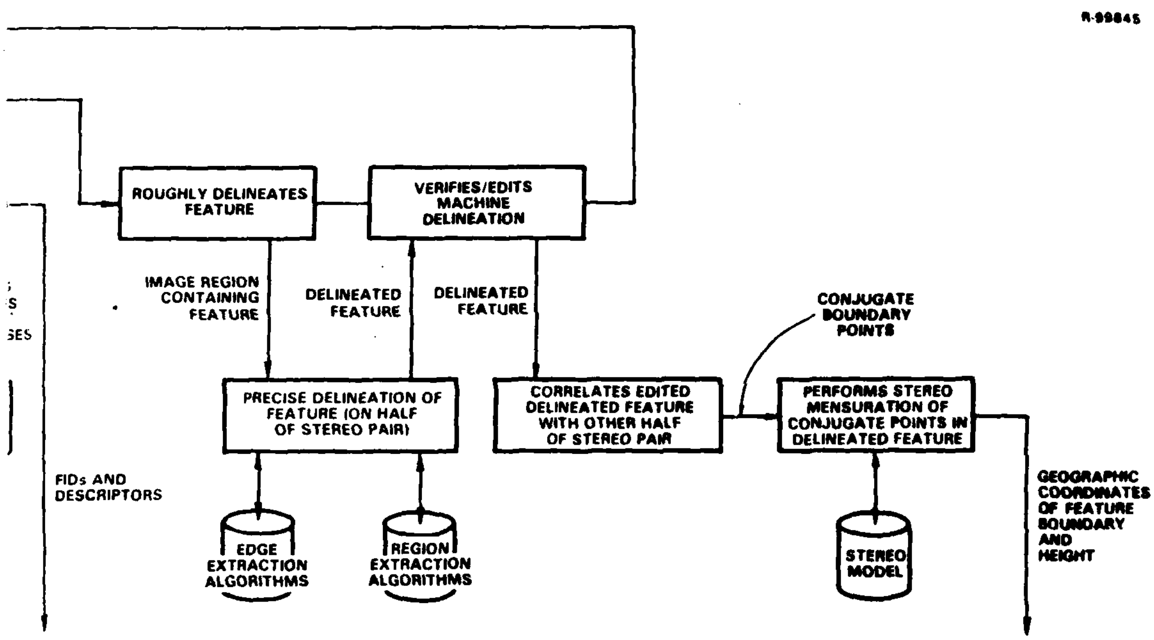


Figure 3.3-1 Refined Feature Identification and Delineation Concept of Operation

2

the analyst in this process, the concept of operation envisions the analyst quickly and roughly delineating each detected feature (e.g., via electronic grease pencil), and passing the delineated feature to the system for more precise delineation. Generalized edge extractors would appear to provide the greatest potential for supporting such a process. The delineated features output from the system would be made available to the analyst in order for him to verify the delineation and edit it as necessary. Having done this, the analyst would resubmit the delineated feature to the system, where correlation of only the delineated feature boundary with its conjugate boundary in the other half of a stereo pair would take place.* The results of this correlation would be fed by the system to a stereo mensuration process which would subsequently derive the geographic coordinates of the boundary. With these coordinates in hand, the analyst would complete his classification/description of the feature.

A general observation concerning this concept of operation is that it partitions the "work" between analyst and machine in such a fashion that they are not only complementary, but are able to be accomplished in a parallel and concurrent fashion. Thus, the efficiency and utilization of both man and machine are maximized. Moreover, the analyst remains in a position to override, modify or correct any results the system might produce. This capability is provided in recognition of the limitations of current machine perception technology. Finally, the concept lends itself in a straightforward fashion to implementation in a digital workstation.

*Recall that we are assuming that mensuration must take place in stereo since no collateral information (e.g., DTED) is available.

3.4 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

3.4.1 Summary

This chapter has outlined a concept of operation for an interactive, semi-automated feature extraction system based on current technology. First, a generic concept of operation for feature extraction was described. Activities within the concept of operation were then identified which appear to be candidates for automation and the application of machine perception. The particular activities selected for automation were feature detection, identification and delineation, and based on the results of Task 3, alternative methods for applying machine perception were proposed. Subsequently, the feasibility issues and cost/benefit trade-offs surrounding the alternative methods proposed were discussed. Based on the results of the latter analysis, a refined concept of operation for a semi-automated feature extraction system was provided.

3.4.2 Conclusions

The conclusions in this chapter are several. With respect to technical feasibility, it was argued that neither source data characteristics (e.g., resolution, scale variation, noise, solar elevation/azimuth) nor digital workstation technology (e.g., storage, processing, communication and displays) posed any feasibility constraints on the implementation of a semi-automated feature extraction system. However, the limitations of current feature extraction techniques and machine perception technology were felt to constitute a relatively high technical risk. In order to reduce this risk, the concept of a minimal attribute set of a feature was defined which would provide necessary but not sufficient information for the

identification of a feature from strictly image-derived attributes. The objective of the MAS was to provide a capability which with high reliability would determine when and where no features of interest were present in an image, and provide cues to the possible location of features where MAS conditions were satisfied. Although the MAS concept substantially improves the potential feasibility of a semi-automated system, whether or not a MAS can be defined for each feature or class of features to be extracted is currently an open problem.

With respect to the cost/benefit tradeoffs associated with the concept of operation proposed, it was argued that in each of the major trade-off categories - machine time vs manual time to perform a given task, man/machine interaction efficiency, equipment vs labor costs, throughput and accuracy - a semi-automated capability was superior to a purely manual capability. However, the tradeoffs can be quantified only in the context of a particular system implementation concept and in conjunction with experiments that would be designed to determine the relative performance of machine perception versus human interpretation in performing selected feature extraction tasks.

3.4.3 Directions for Further Research

Finally, with respect to potentially beneficial areas for experimentation and research, it is recommended that efforts to be devoted to:

- determining the feasibility and defining the characteristics of the minimal attribute set for a number of features of interest

- developing an improved testbed capability for hosting, in a more realistic production environment, experiments to quantify the relative performance of the semi-automated capabilities proposed versus manual capabilities
- defining how the man/machine interface should be developed for this system so that synergistic, rather than conflicting, interaction between man and machine can be realized.

Only after these efforts are completed can the real feasibility and cost/benefit of a semi-automated feature extraction system be determined.

4. REVIEW AND ASSESSMENT OF MULTI-SPECTRAL/MULTI-SOURCE (MS/MS) FEATURE EXTRACTION TECHNIQUES

This chapter reviews and assesses relevant technology for semi-automated feature extraction using multi-spectral (e.g., Landsat Thematic Mapper) and multi-source (i.e., Landsat TM used in conjunction with synthetic aperture radar (SAR) imagery).

In Chapter 2, an assessment of black-and-white (i.e., monoscopic) feature extraction techniques was performed. It was concluded that the computation of physical descriptions from an image is a key step in the feature extraction process, one that must be accomplished before a semantic interpretation can be made. A conclusion was that limited physical information could be gleaned from a single black-and-white image.

As a consequence of the results of Task 3, it became apparent that data from multiple (e.g., stereo), and multi-spectral sensors are required to facilitate the computation of physical descriptions of a scene from an image. For example, relative depth/elevation, and surface orientation can be readily obtained from stereo imagery (Ref. 109). Multi-spectral sensors measure the visible and infrared reflectivity, and thermal emissivity of surface materials, while a SAR provides information on surface roughness and dielectric properties. Separately, and together, they are useful in determining surface material composition.

According to the IU paradigm developed by Kanade (Ref. 3) and adopted by TASC in Task 3 for assessing black-and-

white feature extraction techniques, three distinct levels of information were shown to exist in IU systems: signal, physical, and semantic. Referring to Fig. 4-1, a scene is described by a 2-d array of image intensities at the signal level, by a collection of 3-d objects having specific sizes, shapes, compositions, and relations at the physical level, and by a set of descriptive labels (names or functions) at the semantic level. A signal to physical level transformation attempts to reconstruct some aspect of the physical representation of the scene from the image (in essence trying to invert the image formation process). A physical to semantic level transformation attempts to infer name or functionality from the physical description.

The IU paradigm is applicable to any optical imaging problem for which the goal is to obtain higher-level information from the image. Higher-level information refers both to physical properties of the imaged objects and to the names of those objects. Obtaining the physical properties of an object may be regarded as a measurement process, while identifying its name may be viewed as a recognition process. The IU paradigm

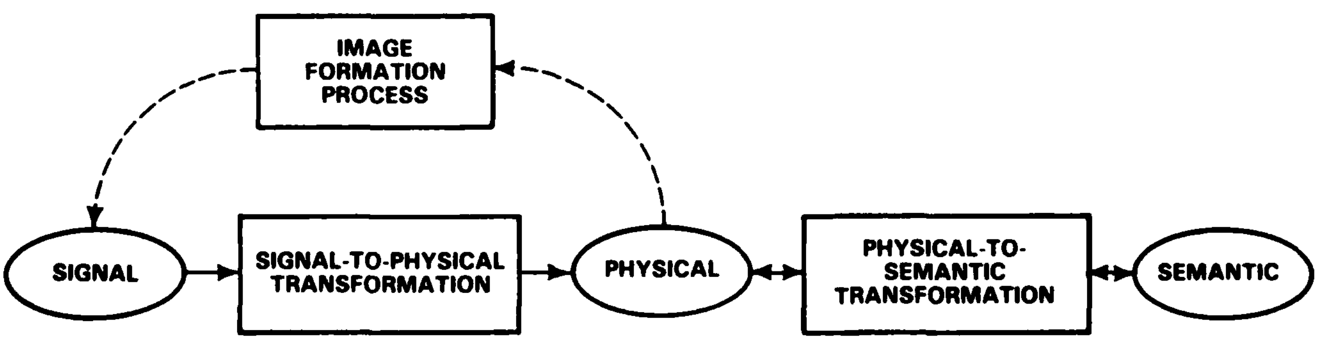


Figure 4-1 Levels of Representation in Image Understanding

shows how semantic categories relate to the physical characteristics of the objects to which they refer, and how semantic information is inferred on the basis of observed physical properties. It secondly shows how the latter physical properties (which have to do with objects and not images) are transformed to imagery, and what is subsequently required to infer object properties from the image.

Use of the IU paradigm serves the following purposes in MS/MS feature extraction:

- It establishes, as a fundamental goal, the determination of the material composition of object surfaces visible in the image from MS/MS imagery
- It suggests a methodology for perceptually organizing the image into regions of the same surface material type, for organizing regions into possible objects based on prior knowledge with regard to the types of materials likely to compose an object, and for identifying objects based on 2-d attributes of regions
- It defines, basic processing requirements in three areas: preprocessing (to condition MS/MS imagery for feature extraction), surface material classification (to infer surface material class based on measurements derived from MS/MS imagery), and object recognition (to group regions of the same surface material type into objects, and to recognize objects as instances of DMA features).

The IU paradigm thus provides a framework or model for assessing candidate MS/MS feature extraction techniques in a meaningful and consistent fashion. The functional model for MS/MS feature extraction shown in Fig. 4-2 consists of three major functional areas:

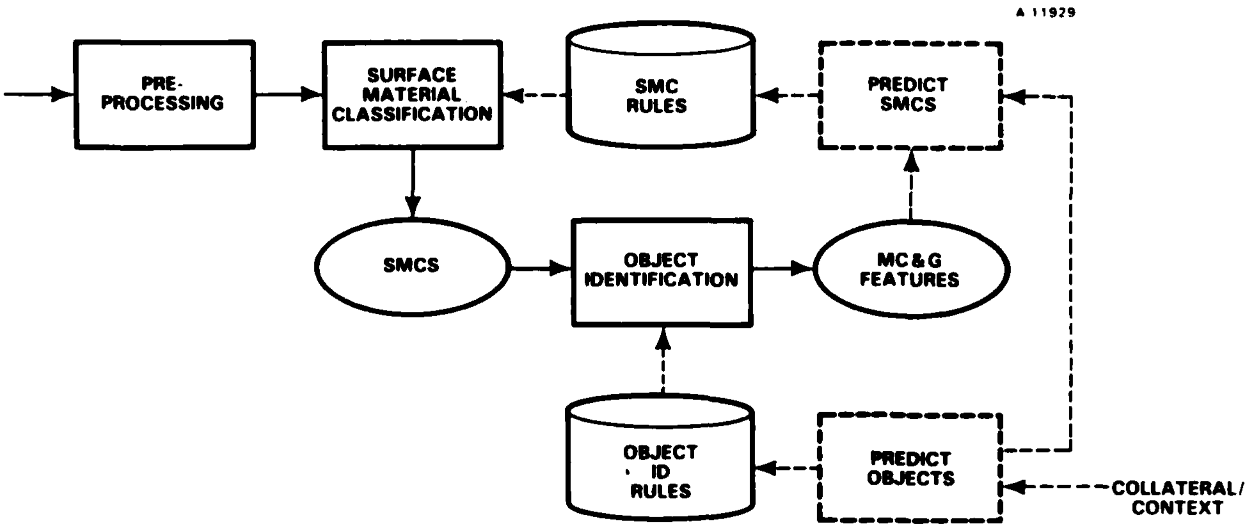


Figure 4-2 Function Model for MS/MS Feature Extraction

- Image processing
- Surface material classification
- Object identification.

The objective of image processing is to normalize data sets acquired at different times, under varying conditions, and by different sensors. It requires image restoration to remove artifacts introduced by the sensor, image registration to bring multiple data sets (e.g., TM and SAR) into alignment, image enhancement for contrast improvement and spatial-frequency sharpening, and image transformation to compute physically-meaningful measurements from the registered, restored, and enhanced data set.

Surface material classification involves determining the composition of object surfaces visible in the image. In terms of the IU paradigm, it is a signal-to-physical level

transformation which infers surface material composition from MS/MS measurements under specified conditions. It requires that once surface materials have been identified, pixels with the same SMC be grouped into connected regions. This represents the organization of a physical description into physically-meaningful units (i.e., regions having similar material properties).

Object identification represents a physical-to-semantic level transformation which groups regions into objects (potential MC&G features) based on prior knowledge about what kinds of regions make up objects, and taken one step further, what kinds of objects can be expected to occur in the scene based on collateral and contextual information. Objects are described in terms of their constituent regions, i.e., by the composition, size and shape of, and the relations between regions and perhaps between other objects as well (Refs. 26, 110).

In this chapter, the above functional model is used to define basic processing requirements for a MS/MS feature extraction system. In particular, image processing must provide adequate information for surface material classification to be carried out, and surface material classification must be able to provide a surface material map of sufficient spectral/spatial resolution and accuracy to support subsequent grouping and identification processes in object recognition. Subsequent sections review candidate MS/MS techniques relative to these requirements.

The remainder of this chapter is organized as follows. Section 4.1 reviews current sensor systems with respect to spatial resolution, spectral bands, and orbital information.

Section 4.2 discusses processing considerations for multi-spectral and SAR imagery. Sections 4.3 through 4.5 reviews and assesses image processing (i.e., registration, restoration, enhancements, and image transformation techniques), segmentation, and classification techniques. A review of applicable computer vision systems/techniques is provided in Section 4.6. Appendices F, G, and H describe new techniques for MS/MS image enhancement, classification, and segmentation.

4.1 REVIEW OF MULTI-SPECTRAL AND SAR SENSORS

This section reviews five imaging sensors for which data is or will soon be available that is of potential use to feature extraction. Three of the sensors (LANDSAT MSS and TM and Spot) are passive electric optical scanners operating in the visible and infrared (IR). The other two sensors (SEASAT SAR, Shuttle Imaging Radar) are synthetic aperture radars (SAR) operating in the L-Band. The characteristics of these sensors and their platforms, together with observations regarding the extraction of information from imagery provided by them are discussed below.

4.1.1 LANDSAT

LANDSAT is a family of spacecraft (Refs. 144, 145) which have carried three different sensing devices: the multispectral scanner (MSS), the panchromatic return-beam vidicon (RBV), and the thematic mapper (TM). There have been five LANDSAT satellites. Table 4.1-1 summarizes the instrument configuration on each. Although the RBV provided superior spatial resolution compared to the MSS, it lacks multispectral capabilities which have limited its use in feature extraction. Hence, it will not be considered further in this report.

TABLE 4.1-1
LANDSAT INSTRUMENT CONFIGURATION

LANDSAT	MSS	RBV	TM
1	√		
2	√		
3	√	√	
4	√		√
5	√		√

In considering applications of LANDSAT data, it must be recognized that LANDSAT-1, 2, and 3, as well as the TM of LANDSAT-4 and 5 have been developmental in nature. The user must be aware of and capable of dealing with sensor noise, gain errors, and uncompensated geometric and band registration errors.

Orbital Considerations - LANDSAT-1, 2, and 3 were all placed in sun-synchronous, near polar orbits. The orbits are essentially circular with altitudes generally in the range of 880 to 940 Km. The 99-degree inclination takes the satellite over all latitudes less than 81 degrees (north and south), permitting the sensing of the entire earth with the minor exception of the extreme polar areas. Motion of the groundtrack is retrograde, each orbit lying to the west of the preceding by 25.8 degrees of longitude. The orbital plane is oriented such that visible sensing can be conducted on the descending mode, which occurs during local mid-morning. The time of equatorial crossing is generally in the range of 8:30 - 9:30 a.m. (local time), depending on the particular satellite. Groundtracks repeat on an 18-day cycle, allowing for the collection of multi-temporal data.

The orbits of LANDSAT-4 and 5 are similar except that the altitude has been reduced to approximately 700 km with a resulting change in the repeat cycle to 16 days. The orbits are again phased to give an equatorial crossing, during the descending mode, at about 9:30 to 9:40 a.m. local time.

Multi-Spectral Scanner (MSS) - The LANDSAT multispectral scanner is the only instrument among those being reviewed to have achieved operational (contrasted with experimental) status. It has been the mainstay of terrestrial remote sensing for more than 12 years. LANDSAT-1 and 2 had an MSS with four visible and near IR spectral bands, designated bands 4 through 7. A thermal IR band (band 8) was added on LANDSAT-3. The bandwidths are shown in Table 4.1-2.

TABLE 4.1-2
LANDSAT-MSS SPECTRAL BANDS

BAND NUMBER	BANDWIDTH (μm)
4	0.5 - 0.6
5	0.6 - 0.7
6	0.7 - 0.8
7	0.8 - 1.1
8*	10.4 - 12.6

*LANDSAT 3 only.

The imagery from the MSS is framed to cover an area approximately 185 km on a side. The instantaneous field-of-view (IFOV) for bands 4-7 is 79 m, but oversampling along with scan line results in a pixel spacing of 56 m. Spacing between scan lines is 79 m. The band 8 IFOV is approximately three times

as large; this data must be resampled to permit co-registration with the visible and near IR imagery.

Thematic Mapper (TM) - LANDSAT-4 and 5 carry the thematic mapper instead of the RBV. This instrument provides data from seven spectral bands in the visible and IR. In addition, it provides improved spatial resolution (30 m, except 120 m for the thermal IR band over earlier LANDSAT vehicles). The TM bands are summarized in Table 4.1-3.

TABLE 4.1-3
LANDSAT-TM SPECTRAL BANDS

BAND NUMBER	BANDWIDTH (μm)
1	0.45 - 0.52
2	0.52 - 0.60
3	0.63 - 0.69
4	0.76 - 0.90
5	1.55 - 1.75
6	10.40 - 12.50
7	2.08 - 2.35

As can be seen from the table, the bandwidth for the TM bands are typically smaller than for the MSS. This improves the ability to discriminate various resources, types of vegetation, and patterns of land use. For example, band 1 is useful for hydrographic studies and the differentiation of coniferous and decidous forests. Band 3 is centered on a chlorophyll absorption band to aid plant species differentiation. Band 4

can detect ferric absorption, an important indicator for some types of mineralization; this band is also important for biomass estimation and the delineation of water. Band 5 can be used to estimate the moisture content in vegetation, and also serves to differentiate between snow and clouds. Band 7 detects a hydroxyl absorption band which is an important indicator of certain clay minerals. A comparison of Tables 4.1-2 and 4.1-3 shows that those portions of the spectrum sensed by TM bands 1, 5, and 7 are not detected by any of the MSS bands. In fact, the interpretation of data from these new bands is a subject of continuing research.

The user of TM data will find that the imagery is formatted in scenes with dimensions identical to those of the MSS. It should be noted that scene centers for LANDSAT-4 and 5 (carrying the TM) do differ from these used with the earlier LANDSATS. This is a consequence of the change in orbital altitude. Also, because of the high spatial data density provided by the TM, quarter scenes (92.5 km on a side) are made available for more convenient data handling.

4.1.2 SPOT

The French are planning to launch their System Probatoire d'Observation de la Terre (SPOT) in the fall of 1985. This satellite will carry two high resolution sensors with a combined field of view of 4.13 degrees. A mirror, steerable by ground command, will permit imaging up to 27 degrees to either side of a given area on several consecutive days. This feature will also permit acquisition of stereoscopic imagery.

SPOT will operate in either of two modes, multi-spectral or panchromatic. The former will provide three spectral bands as indicated in Table 4.1-4. Panchromatic mode will use one broad band.

TABLE 4.1-4
SPOT SPECTRAL BANDS

BAND NUMBER	BANDWIDTH (μm)
1	0.50 - 0.59
2	0.61 - 0.68
3	0.79 - 0.89

While SPOT will not provide the IR coverage of LANDSAT TM, it will give superior spatial resolution. In multi-spectral mode the pixel spacing (at nadir) will be 20 m, while in the panchromatic mode it will be 10 m. Swath width for both will be 60 Km (when centered about the nadir).

The primary advantages of the SPOT sensor are clearly its high resolution and stereoscopic capability. The former will enable the detection of roads, buildings, and other features which are often too small to sense accurately with other devices; the latter may allow improved mapping, terrain classification, and feature identification.

4.1.3 Satellite-Based SAR

Synthetic aperture radars (Refs. 144, 147) have been placed on both the Seasat satellite and on the Space Shuttle (Shuttle Imaging Radar). These sensors were similar in that they were both L-Band radars with a wavelength of 23.5 cm. Resolution was nominally 25 m (with four looks averaged), though pixel spacing in the processed images was about 17 m. The differences between the SARs are discussed below.

SEASAT SAR - The SEASAT SAR collected data for 3 months, late in 1978. Thus only limited amounts of data were acquired. The polarization of this instrument was H-H, that is, horizontal for both transmit and receive. The depression angle of the antenna was approximately 67 degrees, giving incidence angles in the range of 17.5 to 22.5 degrees. Both digital and optical data collection were utilized, depending on telemetry. The digital products have been correlated and, for the most part, are presented in range-doppler coordinates. This results in a significant amount of geometric distortion. Layover effects resulting from terrain relief and the large depression angle are also a major problem in interpreting SEASAT SAR imagery. The digital data is framed as images, 6144×6144 pixels, or approximately 110 km on a side. Dynamic range is generally on the order of four bits.

Shuttle Imaging Radar - The Shuttle Imaging Radar (SIR) was developed as an extension to SEASAT SAR technology. Actually, SIR is a family of instruments: SIR-A, SIR-B and SIR-C. SIR-A was flown on the Space Shuttle in 1981. The instrument characteristics were very similar to Seasat except that the depression angle was significantly reduced to give an incidence angle of approximately 47 degrees. Data recording was restricted to on-board optical methods, thereby mitigating telemetry restrictions.

The next generation system, SIR-B (Ref. 147) is scheduled for launch in the fall of 1984. Many system characteristics remain the same, but the antenna will be moveable to permit variable incidence angles in the range of 15 to 57 degrees. As a result, it will be possible to collect multi-pass coverage for many sites, each pass with a different incidence angle. In addition, the instrument will permit a tradeoff between swath width and number of bits per sample. Up to 6 bits per

sample (bps) is possible, though for some incidence angles this can narrow the swath width to 15 km. In the high bit mode (5 or 6 bps) a calibrator may be used. Uncalibrated (3 to 4 bps) data may be collected in a mapping mode which allows for data collection over a much wider swath. Some squint mode data may also be collected. Digital data will be collected by direct link to TDRS or on an on-board tape recorder for subsequent transfer via TDRS. Optical recording will also be used.

A third instrument, SIR-C, is under development (Ref. 148). This device will offer all the capability of SIR-B but will also offer multiple polarization capability (H-H, V-V, and H-V). Some consideration of a multi-frequency capability for SIR-C is also being given. Launch of this package is tentatively scheduled for the 1988-89 timeframe.

SAR imagery is useful in mapping applications for two reasons. First, because of its longer wavelength (compared to optical or infrared sensors) it penetrates cloud cover without appreciable loss. (Also, since a SAR is an active sensing device, it allows data collection to be performed day and night.) Second, it provides information that is not easily obtained with optical and infrared sensors (e.g., surface roughness, moisture content).

4.2 PROCESSING CONSIDERATIONS FOR MS/MS IMAGERY

In attempting to extract information from multi-spectral/multi-sensor MS/MS data, an analyst is confronted with several sources of potential error: atmospheric degradation, variations in illumination, sensor noise, and sensor calibration errors. All of these effects can contribute to errors in

the estimation of spectral reflectance (Refs. 144, 145). Limitations in spatial resolution and errors in geometric registration can also create problems in attempting to extract linear and point features. This section discusses these factors and data processing methods intended to mitigate their effects.

4.2.1 Optical Data

Optical data refers to the visible and IR data available from multi-spectral sensors, as discussed in Section 4.1. Since SPOT is not yet operational, and the problems in using that data have yet to be determined, the subsequent discussion will be limited to the LANDSAT MSS and TM.

As mentioned above, one source of error is optical data in atmospheric degradation. Problems associated with atmospheric degradation are partially mitigated for LANDSAT by the fact that the sensor views only a few degrees off nadir (this will not be true of SPOT). Nevertheless, the scattering of light by water vapor (haze) can be a problem in some areas. A number of methods for haze correction have been examined (Refs. 144, 145, 149), but only two are in common use. For both, the assumption is made that haze is a simple (constant) additive component to the scene brightness. Its origin is generally high enough in the atmosphere to be independent of surface albedo. Because of the size of water vapor molecules, it is the shorter wavelengths (MSS band 4 and occasionally band 5) that are affected.

In the first method, the magnitude of the haze factor is estimated by examining areas of deep shadow when any perceived illumination is assumed to be due to haze. In the second method, a scattergram of image intensities for the haze

degraded band (e.g., MSS band 4) versus a haze-independent band (e.g., MSS band 7) is plotted on a pixel-by-pixel basis. Assuming that the intensities of the two bands are highly correlated over large areas, a linear relationship is estimated, and the y-intercept for the curve (where band 4 is the ordinate) is the estimate of the haze component.

One must use care in automating the second method since it is important to eliminate pixels for which the intensities for the two bands are not expected to be positively correlated. The assumption may not hold for numerous agricultural crops where plant development may alter the spectral relationships. This is particularly true when one attempts to extend the procedure to use with TM data. Here the narrow bandwidths have been placed to take advantage of various absorption bands (e.g., chlorophyll), thus negating the frequently made assumption regarding band correlation.

It has been recognized from study of airborne thematic mapper data that scattering due to dusts and aerosols can also cause atmospheric degradation of imagery. Problems are most frequent in areas of moderate terrain relief where contaminants are locally trapped in mountain valleys. Because of the lower altitude of dust and aerosol entrapment, the effects are generally albedo-dependent resulting from multiple (Mie) scattering between the polluting particles and the surface. No satisfactory technique has yet been developed for the recognition of these situations without the use of multi-pass imagery. Some work has been done in the application of local area adaptive filters to correct for the effects, once they are detected and quantified. Because of the particle sizes involved, this phenomena effects primarily the 0.8 to 2.2 um bands.

Four types of system-related problems arise with the Landsat sensors. First, with respect to calibration, one must be certain of the calibration standard in use when any particular data set was collected. This is important in using multiple-pass data as the system gain may change from pass to pass. Secondly, regarding the MSS, there is an image striping problem (Ref. 144) which originates with the sensor design. A bank of six detectors are arranged such that together they make a sweep of six scan lines, any given detector being used for every sixth line. Minor bias and differences in response result in the stripping. The problem is not of major consequence until the image is either enhanced for interpretation or used in automated classification. Local area adaptive filtering has been shown to be quite effective in eliminating the problem.

A third problem is related to the striping problem, but is much more defined. The LANDSAT TM has been observed to exhibit several noise problems (Ref. 150), some but not all related to variations in detector response. Adaptive filters are again being investigated as a cure for these problems. Unlike the MSS detector bias problem which is omnipresent, the TM noise problems are time-dependent and much less frequent in occurrence.

Finally, a fourth problem, affecting all TM data, arises from the fact that there are three different focal planes in the sensor, one for the visible and near IR, one for the short wave IR (the 1.6 and 2.2 um bands) and one for the thermal IR. Analysis has demonstrated some significant misregistration of data from the separate focal planes (Refs. 150, 151, 152). A simple geometric translation during the data preprocessing stage is all that is required, but failure to due so may well cause misidentification of very narrow features or of those pixels bounding larger domains.

In addition to the corrective preprocessing discussed above for atmospheric degradation of system noise, techniques for handling variations in scene illumination may also be required. Differences in solar azimuth and elevation (for multi-scene evaluation) as well as local variations in illumination due to terrain relief (including shadowing) can have significant effects on the apparent spectral reflectivities in a scene. Attenuation of sunlight by cloud cover (not necessarily within the subscene of interest) may also be important. All of these effects may in large measure be mitigated by using band ratio images for analysis of the scene. By dividing the perceived (detected) radiance of one band by that for another, one is essentially bringing out anomalies in the spectral reflectance in a way which "normalizes" them to the available (local) scene illumination. Advantageous scaling of band ratio images is, however, achieved only with considerable experience. In principle, one might perform terrain corrections through use of digitized, co-registered topographic data. In practice, however, this approach is computationally expensive, and topographic data of sufficient accuracy and spatial resolution is not generally available.

4.2.2 SAR Data

Preprocessing of synthetic aperture radar (SAR) data (assuming the signal data have already been correlated to form an image) is primarily directed towards geometric corrections (Refs. 144, 153). SAR images are most commonly formatted in range-doppler coordinates. For radar illumination perpendicular to the orbital velocity vector, the primary correction is a conversion of slant range to ground range. Earth rotational velocities will, however, contribute to some error in the doppler range coordinate. Warping or "rubber sheeting", that is, fitting the data to some higher order polynomial surface, is

commonly used to make the necessary geometric adjustments, but this process requires the recognition of ground control points. Depending on the type of terrain and the radar incidence angle, this can be a very difficult task for some scenes using L-Band data. The data user must be cautioned that residual errors on the order of 10 to 15 pixels (or greater) may be introduced. Use of an accurate SAR system model coupled with accurate spacecraft ephemerides would enable improved registration, but ephemerides of sufficient accuracy are often unavailable.

A second type of registration error arises from terrain relief. The cross-track dimension in SAR is derived directly from timing of the radar pulse to determine the range to target. By using a flat (or other specified) model for the sensed surface, the radar return at any particular time is correlated with the terrain at a specific range. If the terrain is in fact sloped perpendicular to the "sighting" vector, there is in reality a much greater area at the same radar range. This creates an ambiguity, referred to as layover, in the processed scene. Its occurrence is more frequent at small incidence (large depression) angles. Some techniques for dealing with layover through the use of digitized topography are being studied. At present, however, preprocessing for this and related terrain effects, including shadowing, is limited to recognition of the condition and use of higher-order geometric correction to mitigate some of the resulting distortions.

Calibration of radar data must be considered for multiscene studies. Both the SEASAT SAR and SIR-A were uncalibrated devices. To correlate data from multiple scenes, backscatter adjustment made from the data implicit in the imagery is necessary. Some of the SIR-B data will be processed using an internal calibrator signal. In analyzing the data one must always recognize the limited dynamic range of the radar system and understand the antenna radiation pattern so as to relate properly

the backscattered signal to the derived (and arbitrary spaced) pixels. Depending on the number of looks averaged in processing the SAR data (four looks are commonly used), the resultant image may have a considerable amount of "speckle". This high frequency noise is generally not problematic for human interpretation, but may frequently result in large errors when machine interpretation or classification is performed. These errors may be substantially reduced by using a median or small area averaging filter (typically 3 × 3 pixels), but loss of information regarding small scale features (e.g., roads and buildings) is likely to occur. A two-phase process, where the unfiltered data is used to evaluate selected smaller features followed by use of filtered data to evaluate larger area features, may be beneficial.

4.3 IMAGE PROCESSING

In an MS/MS feature extraction system, image processing techniques are used to normalize (i.e., register to a common coordinate system and scale) images acquired at different times, under varying conditions, and by different sensors, and to restore/enhance each image prior to surface material classification and feature extraction. Image processing includes such functions as restoration techniques to recover data that have been lost or degraded due to sensor dropouts or noise, and registration techniques to register different data sets to one another or a common coordinate system. Image processing functions also include such techniques as color (multi-spectral) and spatial multi-band enhancements. Color enhancements include image transformation techniques that are useful for projecting multi-dimensional data into physically-meaningful coordinate systems, while spatial multi-band enhancements use information from multiple spectral bands/sensors to enhance an image.

4.3.1 Restoration Techniques

Restoration techniques are used to remove noise from imagery data so that subsequent registration, enhancement, and analysis can be performed on images with a high signal-to-noise ratio. The precise separation of any noise from image data must be based on quantifiable characteristics of the noise signal that distinguish it uniquely from the other image components. In addition, the processing of the noise must be done in a manner that minimizes the distortion of the desired image data. For typical Thematic Mapper (TM) or Multi-Spectral Scanner (MSS) data, the main types of noise are periodic noise (coherent instrumentation noise), striping (sensor gain variations), and data dropouts or spikes (isolated data disturbances). For SAR imagery, the primary types of noise are data line dropouts and peak saturation.

Periodic noise within multi-spectral imagery may be caused by the coupling of periodic signals related to the scanning instrumentation into the imaging electronics. The recorded images contain periodic interference patterns, with varying amplitude, frequency, and phase superimposed over the scenes of interest. In areas of scenes which are extremely uniform (e.g., over bodies of water), system noise can be made apparent by contrast stretching. For typical spaceborne sensors, the phase coherence time period (for noise) is long compared to frame acquisition times. Therefore, the periodic noise appears as a regular two-dimensional pattern. If one examines the magnitude of the FFT of the noisy image, then one can detect the regular periodic noise components as peaks in the transform domain.

Multiple frequency bands from a multi-spectral scanner can be used together to yield a better estimate of the common

mode noise spectrum, assuming that the images for each band are acquired simultaneously by the same sensor system. For a multi-band system, systematic frequency noise can be estimated from the coherent component across bands. Also, model-based spectrum estimation techniques can be used to improve the estimate of the periodic noise spectrum (Ref. 111). These latter methods are ideal for detecting spatial sinusoids in random data.

Having estimated a noise spectrum, frequency domain notch filtering can suppress systematic noise if the data and noise spectrum peaks are well separated. If the data and noise spectrum overlap significantly, then the image data can be degraded after filtering. An alternative to modifying the composite (noise plus data) spectrum by notch filtering is to interpolate the magnitude of the spectrum across the noise peaks (Ref. 112). This avoids the problem of estimating the attenuation factor required by the notch filter.

An adaptive scheme for eliminating high frequency system noise from the data is based on phase-locking a noise reference signal (single frequency sine wave) in the spatial domain and subtracting it directly from the image data (Ref. 113). This method works well in areas of low contrast (low spatial detail) such as over water or barren fields, since the phase-lock can lock onto image detail rather than the noise signal. An improvement over the spatial method described above can be made by phase-locking the noise reference to the coherent part of the multi-band data set, which is due partly to the common mode noise signal.

Regular striping may occur in images taken by multi-detector sensors. Multiple detectors are used (for each frequency band) to image a group of lines during one mirror sweep. For example, six detectors are used for each band of the MSS.

The TM contains a total of 100 detectors (16 for each of the 6 visual and infrared bands, and 4 for the thermal band). Striping occurs in uncalibrated data because the individual detectors exhibit slightly different gain and offset variations. Striping frequently occurs in "ground processed" (presumably calibrated) data also, where although it is much reduced, it is not completely eliminated. The reason for this is that the striping is digitized between the time the radiance is sensed (on the satellite) and corrected on the ground. This radiance quantization causes a loss of precision which leads to the observed stripes.

A technique to correct for regular striping involves matching the first order intensity statistics for each group of parallel lines imaged by a particular detector. For each detector group, a gain and offset are computed to match the current line group with the overall image statistics. This linear correction is particularly data dependent in its effect, and although providing a global improvement, it may introduce artifacts in detail (Ref. 114). Nonlinear sensor effects distort the intensity distributions sufficiently so that the linear correction does not eliminate striping completely. A nonlinear correction (histogram remapping), obtained by matching the cumulative histogram of the individual detector line groups to the cumulative histogram of the entire image, successfully reduces striping (Refs. 115, 116). Another interesting technique uses a probabilistic approach to remap intensity values so that the total accumulated error approaches zero (Ref. 113). Since the output imagery data is usually integer, a probabilistic approach allows remapped values to be real, on the average over the whole image.

Current destriping techniques do not take full advantage of multi-band data. The local correlation property between

bands can possible be used to estimate what the striped imagery should look like and then that estimate can be used to compute correction gains and offsets or remapping histograms. This technique relies on the fact that striping is not typically apparent on all bands of multi-band imagery.

Data dropout or spike noise within multi-spectral imagery can be caused by errors in data transmission or by temporary disturbances in the analog electronics. These noise patterns manifest themselves as isolated line segments or isolated pixels that deviate significantly from their surrounding data. A simple method for detecting and filtering spike noise involves comparing each pixel with its immediate neighbors. If all differences exceed a certain threshold, the pixel is considered a noise point and is replaced by the average or median of its neighbors. A similar scheme can be implemented for removing line segment noise.

Since data dropout and spike noise are generally uncorrelated across bands in multi-band imagery, a multi-band approach can be effective in detecting noise segment and points. By examining the local correlation (which should always be high), noise points will show up as points of low local correlation. Alternatively, if a two-dimensional linear prediction is performed across bands, noise points can be detected by looking at the error residuals (Ref. 117).

For SAR data, systematic radiometric corrections are useful for shading corrections both along and across track. A major difficulty with SEASAT SAR data, for example, is caused by the limited dynamic range in many parts of the system (Ref. 118). A problem with calibration pulses exceeding the dynamic range of the data link results in white streaks in the imagery. Another result of signal saturation is weak-signal

suppression. In this case, the signal from a dim target is suppressed by a very bright target in proximity. The processing of these two types of artifacts will require some type of "intelligent" detection and filtering.

4.3.2 Rectification and Registration Techniques

This section describes a class of techniques used to geometrically transform imagery data to a selected coordinate system. Two important classes of mappings are transformations that relate image coordinates (x,y) to a geodetic coordinate system, and transformations that relate the coordinate systems of two different images. Precise geodetic coordinates associated with image pixels are required to produce cartographic projections of images for the joint analysis of map and imagery data. Coordinate transformations between images are employed in the relative registration of several images of the same scene (e.g., multi-spectral, multi-temporal, or multi-source images) so that multi-dimensional processing can be performed pixel by pixel on the combined data set.

The most direct method for rectifying or registering digital image data is by means of polynomial remapping (also called "rubber-sheeting" or "warping"). For image to map rectification, UTM map coordinates and image (pixel, scan-line) coordinates are computed for selected ground control points (GCPs). These GCPs then define (in a least-squares sense) the polynomial used for remapping. For image-to-image registration, the GCPs are defined in both image coordinate systems. The minimum number of GCPs required to uniquely specify the remapping polynomial is dependent on the degree of the polynomial used. For example, first, second, third, fourth, and fifth degree polynomials require a minimum of 3, 6, 10, 15, and 21 GCPs, respectively. One such program package which used this

approach was the Digital Rectification System (DIRS) (Ref. 119) used by NASA for the rectification of early LANDSAT MSS data.

A second approach for image rectification is the systematic approach primarily used for SAR data. In the systematic approach, the predictable errors are identified and correction terms are generated based on geometric parameters. The algorithm for removal of these geometric distortions are derived from an understanding of the entire radar imaging system (from the radar instrument and its spacecraft platform through to data processing to an interpretable image). The predominant distortions found in SAR imagery are along-track (azimuth) skew and ground range nonlinearity (unequal scaling) in the along-track and cross-track directions. In implementing this approach to rectification, coordinates of control points from both the SAR image and a reference image/map are first selected. Then rotation angle, range scale, track scale, and skew angle corrections are computed systematically from the coordinate data using least-squares analysis. A functional description of this approach can be found in Ref. 120. The systematic approach does not correct nonlinearities in the SAR image that might be caused by terrain variations (e.g., layover) or SAR platform variations.

A summary of the cartographic accuracies that can be achieved with Landsat and Seasat SAR data are tabulated in Table 4.3-1. In general, due to improvements in the pointing and attitude control of the spacecraft, and in the ground processing procedures, the cartographic quality of the LANDSAT-4 data is significantly better than that from LANDSAT-1, 2, and 3. For example, LANDSAT-1, 2, and 3 data were acquired from a platform meeting specifications for a 0.7 degree pointing accuracy and an attitude stability of 10^{-2} degrees per second. Overall, the practical accuracies of both LANDSAT-4 MSS and TM data sets are limited to about ± 1 pixel for subscene and whole scene

TABLE 4.3-1
ACHIEVABLE REGISTRATION ACCURACIES

SENSOR	LANDSAT 1,2,3 MSS	LANDSAT-4 MSS	THEMATIC MAPPER	SEASAT SAR
IFOV	79 m	83 m	28.5 m	25.4 m
Remapping Function	5th order poly.	3rd order poly.	1st or 2nd order poly.	Systematic
GCPs Required	dense set	20-30	5-10	20-40
Misregistration Error	±0.63 pixels ±50 m	±0.66 pixels ±55 m	±0.70 pixels ±20 m	±2 pixels ±50 m

areas. Three factors limit geometric accuracy. The first is image resolution which limits the location of GCPs to about a half pixel. A second factor is map and digitizing errors which average about 15 m, while the third factor is terrain relief, which for moderate terrain has been shown to produce displacements between 10 to 30 meters (Ref. 121). The effects of exaggerated relief can be greatly reduced by selecting GCPs at midrange elevations.

4.3.3 Enhancement Techniques

The goal of image enhancement is to aid the photointerpreter in the execution of previously mentioned preprocessing functions and in the extraction and interpretation of information from the data. Enhancement methods may be divided into the following categories:

- Contrast enhancement

- Spatial enhancement
- Radiometric multi-band (color) enhancement
- Spatial multi-band enhancement.

The first two categories of enhancement are applied to individual components of multi-band imagery; the utility of these techniques for enhancing monochrome imagery is described in Appendix A. In the multi-band scenario, contrast and spatial enhancement techniques improve the visibility of ground features, allowing ground control points for registration to be easily identified.

In particular, color enhancement techniques involve linear and nonlinear combinations of component images to produce physically meaningful imagery or pseudo-color composites. Spatial multi-band enhancements are a relatively new class of enhancements that incorporate information from different spectral bands or other sources to enhance a given image in terms of spatial resolution or noise content.

The color enhancement techniques that were assessed include ratioing, principal components, canonic correlation, and the tasseled cap transform. Multi-spectral images may be enhanced by ratioing individual spectral components and then displaying the various ratios individually or as color composites. The technique of ratioing consists of forming a new image from a given pair of images by dividing the first image by the second on a point-by-point basis. It provides relative information and reduces the effects of uneven illumination. The ratio is a measure of "color", (i.e., the relative weight of one band to another). Ratioing has been a useful processing technique for geologic applications (e.g., to enhance the differences between mineral types) but is not commonly used

elsewhere (e.g., land use determination). Preprocessing functions for noise removal, radiometric distortion, and spatial distortion must be applied first to the data, since ratioing exaggerates anomalies. Since a ratio of two positive numbers falls in the range zero to infinity, it is necessary to reduce the dynamic range inherent in this technique. Various remapping functions are used, such as arctangent, logarithm, and cube-root. The merits of these remapping functions are reviewed in detail in (Ref. 122).

The selection of the most useful ratios and their combinations into color composites continues to present a problem. The number of possible ratios from a multi-band image with P components is P(P-1). In order to effectively use the ratio technique, one must have a priori knowledge of the useful ratios for the types of features that are being enhanced (e.g., *MC&G features*). For some basic features, such as water/land boundaries, the ratio of TM band 1/band 4 can be used. For other features, appropriate ratios remain to be determined.

Principal component transformation (alternatively called the Karhunen-Loeve (K-L) transform) of multi-band data involves computing a new set of component images that are uncorrelated and are ranked so that each component has less overall variance than the previous component. In typical applications, principal component images are individually enhanced and combined in various combinations to produce false-color composite images. Since multi-spectral images often exhibit high correlations between spectral bands, the principal component transformation can be useful since it reduces the redundancy in the data.

The procedure for using the K-L transform for enhancing multi-spectral consists of two major steps:

- Computing the covariance matrix of the multi-spectral and its eigenvectors in the spectral dimension
- Transforming the given image vector to principal components by multiplying it by the appropriate eigenvector.

For an image that contains agricultural and urban areas, the first principal component appears to be correlated with vegetation and crops. The second component is correlated with bare soil areas, while the third component is indicative of urban and manmade areas, such as buildings and highways. On the other hand, for images that contain large bodies of water, suburban and urban areas, the first principal component provides a large contrast between land and water (Ref. 113), while the second component provides good cultural feature definition (roads and buildings). Clearly, the correlation of principal component to land feature is highly dependent upon the content of the source imagery.

A transformation that is not an enhancement in a strict sense but aids in the interpretation of data is provided by canonical correlation (Ref 162). Canonical correlation analysis attempts to derive a linear combination from each of two sets of image data in such a way that the correlation between the two linear combinations is maximized. Several pairs of linear combinations (termed canonical variates) can be derived. Canonical correlation can be applied to multi-spectral analysis by grouping the input data as one set of data and known classifications (i.e., soils, vegetation, urban, etc) as the other set of data. In this way, linear combinations of the bands of input data can be generated which are maximally correlated with certain

groups of classifications. These canonical variates are essentially equivalent to the principal components produced by principal-component analysis, with the exception that the criterion for their selection is different. Whereas both techniques produce linear combinations of the original image bands, canonical correlation analysis tries to maximize the relationship between two sets of images instead of accounting for as much variance as possible within one set of images.

The Tasseled Cap transformation (Ref. 123) has been widely used to transform MSS data, and to a lesser extent TM data, to physically-meaningful measures such as soil brightness, and vegetative cover. Experience has shown that the data variability for the MSS four-dimensional data set is largely confined to a single plane for agricultural regions (alternatively, the first two principal components account for most of the variance in the data). The Tasseled Cap transformation performs a linear transformation (rotation) on the data such that a head-on view of the data variability plane is achieved.

The appropriate matrices for MSS and TM data transformations are listed in Tables 4.3-2, 4.3-3, and 4.3-4. The first two axes are called the "brightness" and "greenness" features, which can be readily associated with physical properties of the scenes. The third axis can be associated with wetness (turbid water) or dryness (concrete or urban). As more data from the TM and other sources become available, representing a broader range of cover classes, new coefficients are likely to be developed.

The second major category of multi-band enhancements are spatial enhancements which include thermal band sharpening and SAR smoothing. These techniques use local correlation

TABLE 4.3-2

LANDSAT-2 MSS TASSELLED CAP
TRANSFORM COEFFICIENTS

FEATURE	BAND 1	BAND 2	BAND 3	BAND 4
Brightness	0.33231	0.60316	0.67581	0.26278
Greenness	-0.28317	-0.66006	0.57735	0.38833
Third	-0.89952	0.42830	0.07592	-0.04080
Fourth	-0.01594	0.13068	-0.45187	0.88232

TABLE 4.3-3

LANDSAT-4 MSS TASSELLED CAP
TRANSFORM COEFFICIENTS

FEATURE	BAND 1	BAND 2	BAND 3	BAND 4
Brightness	0.37821	0.58460	0.69311	0.18655
Greenness	-0.31681	-0.63852	0.62665	0.31501
Third	-0.86920	0.49080	0.05997	0.00138
Fourth	0.03272	0.09822	-0.35115	0.93057

TABLE 4.3-4

LANDSAT-4 TM TASSELLED CAP TRANSFORM
COEFFICIENTS FOR REFLECTIVE BANDS

FEATURE	BAND 1	BAND 2	BAND 3	BAND 4	BAND 5	BAND 7
Brightness	0.33183	0.33121	0.55177	0.42514	0.48087	0.25252
Greenness	-0.24717	-0.16263	-0.40639	0.85468	0.05493	-0.11749
Third	0.13929	0.22490	0.40359	0.25178	-0.70133	-0.45732
Fourth	-0.83104	0.07447	0.42144	-0.07579	0.23819	-0.25247
Fifth	-0.32530	0.05361	0.11485	0.11140	-0.46571	0.80549
Sixth	0.11381	-0.80714	0.42038	0.06686	-0.01629	0.02706

properties to enhance or filter one band relative to all other bands. The thermal band sharpening technique uses all the visible and IR TM bands to optimally sharpen (using least squares) the thermal band which is one-fourth the resolution of the other bands. The Seasat SAR smoothing technique uses pre-registered TM bands to optimally smooth noisy SAR data. These two techniques are described in detail in Appendix F.

4.3.4 Summary

Table 4.3-5 summarizes representative techniques from each major category of image processing techniques discussed in this chapter. For each technique identified, a brief description of the technique and its utility to multi-spectral or multi-source image processing is provided.

4.4 IMAGE SEGMENTATION

The objective of image segmentation is to group an image into physically-meaningful edges or regions prior to classification and subsequent interpretation. This chapter assesses a class of image segmentation techniques that are useful for partitioning an image into homogeneous regions. Such techniques are found in computer vision systems such as the prototype system developed at Kyoto University for interpreting aerial photographs (Ref. 110) and in VISIONS developed at the University of Massachusetts (Ref. 26).

Region segmentation techniques may be divided into three classes: those which use local spatial criteria for region growing, those which use the distributions in feature

TABLE 4.3-5
SUMMARY OF MS/MS IMAGE PROCESSING TECHNIQUES

CLASS	TECHNIQUE	DESCRIPTION	UTILITY
Restoration	Frequency-domain notch filtering	Locate peaks due to noise in the frequency-domain and apply notch or interpolating filter polating filter	Removal of periodic noise (e.g., sinusoids) in image
	De-striping (histogram remapping)	For each line in the image, match histogram to the histogram of neighboring detector line groups by non-linear remapping of pixel intensities on a band-by-band basis	Removes regular striping (fixed pattern noise) due to gain/offset differences between detectors
	Dropout detection/restoration	Detect dropouts as points of low correlation between bands; replace by the average/median value of neighboring pixels	Restores data dropouts and removes spike noise
Rectification/Registration	Polynomial remapping	Specify ground control points (GCPs) to define (in a least-squares sense) a polynomial to map pixels in one image to corresponding pixels in the other	Relative registration of MS/MS data sets prior to multi-dimensional processing (e.g., classification)
	Systematic approach	Similar to above except that aspects of the overall system (sensor, platform, data processing) are taken into account	Registration of data set to a ground coordinate system (e.g., UTM)
Enhancement/Image Transformation	Ratios	Divides two images on a pixel-by-pixel basis	Reduces the effects of uneven illumination (shadows); enhances geologic structures
	Principal components transformation	Computes a set of uncorrelated images ranked in order of de- of decreasing variance; the components having the highest variance are then selected	Used to reduce the dimensionality of the data prior to classification
	Canonical variates/Tasseled cap transformation	Linear multi-spectral data transformations which produce scene-independent measurements of physically-significant quantities	Provides physically-meaningful measures such as greenness or wetness from multi-spectral data
	Multi-band enhancement	Uses local correlation properties to enhance or filter multi-spectral/multi-band imagery	Thermal band sharpening, SAR smoothing (speckle reduction)

space* for region splitting, and hybrid techniques which use both local spatial information and global feature information. Region-growing and hybrid techniques are treated in Section 4.5 in the context of region classification. This chapter focuses on the use feature-space techniques, in particular, data clustering, recursive region splitting, and mixture modeling for identifying clusters (or modes) in the data prior to classification.

4.4.1 Image Segmentation by Clustering

Coleman (Ref. 31) proposed an unsupervised learning approach to segmentation based on data clustering in n-dimensional spaces. Since the number of clusters (i.e., the number of distinct regions in the image) is not known a priori, the algorithm starts at k=2 clusters and iteratively assigns each sample (pixel) to the nearest cluster. When the cluster means converge, the product of the within-cluster and between-cluster scatter is computed. If it has not decreased, k is incremented by one and the clustering is repeated. The best number of clusters is the value of k which maximizes the above quality measure.

For large multi-dimensional images having many clusters, data clustering is an extremely time-consuming solution to image segmentation. If the clusters are close together, the convergence rate will be slow. Also, the solution may only be globally optimal, thus requiring many runs of the algorithm to ensure that the correct solution has been obtained (Ref. 124). On the other hand, clustering is a general and powerful data analysis technique in that it processes all dimensions at once, rather than one feature at a time (e.g., as is done in histogram splitting).

*In this context, the multi-dimensional space spanned by registered multi-spectral/source vector data

4.4.2 Thresholding and Region Splitting

The idea of computing thresholds from histograms for image segmentation was first suggested by Prewitt and Mendelsohn (Ref. 125). Ohlander (Ref. 30) developed a recursive region splitting technique in which regions are split repeatedly into smaller regions until all regions are unimodal, In each region one-dimensional histograms of the red, green, blue, intensity, hue, and saturation are computed. The algorithm selects the histogram with the best peak definition and uses it to compute a threshold for splitting. The idea is that splitting large regions should help in extracting the smaller regions whose distributions tend to be obscured by those of the larger regions.

Histogram-splitting will only be successful if at least one of the histograms in the current region has well-defined peaks. The use of redundant image features (i.e., intensity, hue, and saturation in addition to red, green, and blue) helps to insure this; thus, the features should have some degree of correlation. (On the other hand for data clustering, uncorrelated features as provided by the Karhunen-Loeve transformation should be used.)

4.4.3 Mixture Model Approaches

The use of mixture models for segmenting medical images was first suggested by Chow and Kaneko (Ref. 99). They computed histograms over small overlapping blocks in chest radiographs in order to compute local thresholds for extracting lung boundaries. If the histogram passed a bimodality test, it was decomposed into a mixture of two normal densities. Mixture parameters were determined by curve-fitting and used to compute a maximum likelihood threshold. More recently, in the remote sensing community, a mixture model approach, used

to estimate crop areas, has been developed by Lennington (Ref. 126). The fitting of a mixture model to the observed probability density is accomplished using an algorithm called CLASSY. CLASSY assumes that the mixture components are multivariate normal densities. The number of components is estimated via a sequence of hypothesis tests using a likelihood ratio criterion. The parameters of each component are estimated using the iterative fixed point equations resultings from a maximum likelihood formulation.

Another mixture decomposition algorithm based on an analysis of zero-crossings in the second derivative of histograms is currently under development at TASC. The analysis is performed at various scales or resolutions. The technique computes the number of modes (i.e., the number of component densities, assumed to be Gaussian in form), and estimates the parameters of the component densities. The technique is described in detail in Appendix H.

4.4.4 Summary

Table 4.4-1 summarizes the above segmentation techniques. In feature space approaches, the underlying assumption is that each kind of region in the image gives rise to an associated distribution(s) in feature space. In reality, however, distributions from different kinds of regions often overlap (e.g., in one dimension, the histogram may not exhibit well-defined peaks). Thus, techniques which rely on the presence and/or formation of well-defined peaks in histograms may be unable to split an image into regions in some situations.

The mixture model approach to image segmentation is attractive for two reasons. First, it estimates the underlying probability densities and uses them as the basis for

TABLE 4.4-1
SUMMARY OF REGION SEGMENTATION TECHNIQUES

TECHNIQUE	DESCRIPTION	UTILITY
Clustering	Iteratively assigns pixels to groups or clusters based on their relative proximity in feature space	Useful in segmenting multi-dimensional imagery
Region Splitting	Recursively splits regions until all regions are unimodal	Useful in segmenting color imagery
Mixture Modeling Approach	Estimates the component probability densities which constitute the histogram; segment by thresholding	An analysis technique for use in labeling the various modes in a histogram (e.g., shadows, speculars)

segmentation. Second, once the class-conditional probability functions are known, a posteriori probabilities can be computed and updated locally using either Markov models or spatial relaxation techniques (discussed in the next chapter). The updated probabilities may then be used to classify the pixel, or may be used as a measure of the typicality of a pixel relative to a particular class. This would allow semantic information to be factored into the latter stages of the classification process to resolve ambiguity using, for example, contextual information (Ref. 157).

4.5 IMAGE CLASSIFICATION

In the MS/MS feature extraction system, image classification is the process of labeling image pixels or regions as instances of surface material types using spectral information alone, or in conjunction with spatial and temporal information.

The following sections describe classification techniques which use spectral information on a point-by-point basis only (pixel classifiers), and techniques which use local information as well (region classifiers). The use of multi-temporal data and vegetative growth models for classifying agriculture is reviewed, and the problem of signal variability between scenes and the use of signature extension algorithms to map signatures (material types) in one image to those in another is discussed. Finally, a knowledge-based approach for representing and classifying material types is outlined.

4.5.1 Pixel Classification

Pixel classifiers use the intensities of spectral bands and transformations of spectral bands on a pixel-by-pixel basis as features for classification. Surface material classes are represented in terms of either class-conditional probability densities (non-parametric approach), or the class statistics of an assumed class-conditional probability density (parametric approach). Most techniques are straightforward applications of decision-theoretic pattern classification theory (see Swain (Ref. 127), for example).

For pixel classifiers, prototypical or training region(s) must first be selected to train the classifier. Class-conditional probability densities are obtained empirically by computing histograms over training regions for each class. Alternatively, if one assumes a particular functional form for the class-conditional densities (e.g., multi-variate Gaussian), the parameters which characterize each class (the relative frequency, mean vector, and covariance matrix for the classes) are estimated in the training region.

In estimating class statistics, one must be aware of sample size considerations (i.e., the minumum size of a training region). Foley (Ref. 128) has shown that for the two-class problem with multi-variate Gaussian densities, the ratio of the sample size to the number of dimensions should be greater than three. This is a conservative lower bound especially when the form of the density is not known.

Accurate training is predicated on the availability of ground truth, or on an image analyst's ability to infer ground truth from collateral data sources (maps and charts) or directly from the imagery. In some DMA applications accurate ground truth may not be available, or it may be outdated. Having to rely on the image analyst to provide ground truth is neither practical nor wise since he may not be familar enough with the sensor or the scene to be able to infer ground truth and may provide incorrect information based on subjective judgements. Therefore, problems may exist in using statistical classification techniques for feature extraction.

The classification phase involves using the information obtained by training on regions of known material type in one image to classify unknown regions in the same or in another image. A commonly used classification strategy is to select the class having the largest a posteriori probability of occurrence. Such an approach minimizes the probability of mis-classification. When the classes are (or may be assumed to be) equally likely, the Mahalanobis distance (Ref. 158) may be used as a similarity measure. The classification rule then is to pick the class with the smallest Mahalanobis distance. If the spectral features are uncorrelated (after principle components transformation, for example) and have identical variance, a nearest neighbor (or minimum Euclidean distance) classification rule may be used. The minimum probability of error classifier is

the most costly to apply (in terms of computational expense), followed in cost by the Mahalanobis and minimum distance classifiers. In any case, the performance (i.e., the probability of misclassification) is determined by the statistical separability of the surface material classes.

4.5.2 Region Classification

Due to potential overlap in the tails of the class-conditional probability densities, a classifier which decides a pixel's class on the basis of the spectral signature at a single point in the image can introduce error. Landgrebe (Ref. 129) describes four classes of techniques which can be used to exploit the dependence between adjacent pixels to improve classification accuracy: those which use textural information (e.g., the grey-tone spatial dependence matrix), those which use local spatial information (i.e., region-growing followed by sample classification), those which use contextual information, and those which use relaxation labeling techniques. The use of texture and relaxation techniques were addressed in Chapter 2. Techniques which use local spatial information and context are described below.

The classification of multi-spectral data through the Extraction and Classification of Homogeneous regions (ECHO) is a two-step process which "grows" spectrally homogeneous regions, and classifies them on the basis of their sample distributions (Ref. 130). It uses a likelihood ratio test to decide if adjacent regions are similar on the basis of their probability densities (assumed to be Gaussian). The technique suffers from the problem of not having a sufficient number of samples in the early stages to estimate the probability densities reliably, e.g., when the image consist almost entirely of atomic (usually 4 pixel) regions. One solution to this problem

is to reduce the dimensionality of the data (i.e., reduce the number of spectral bands), so that smaller sample sizes can be tolerated.

The classification error associated with the above technique is shown to be dependent on the annexation threshold. For very small thresholds few regions form, and the classification error equals that of the pixel classifier since little or no annexation takes place. As the threshold increases, the statistical test becomes less stringent, a greater amount of inhomogeneity is tolerated, and larger regions form. Classification accuracy increases as the annexation threshold increases to a point, and then decreases. (Improvements in accuracy of the order of 3% are reported in Ref. 130). While threshold selection is somewhat ad hoc, the ability to control the sensitivity may be considered a desirable feature in an interactive feature extraction system.

Chittineni (Ref. 131) has developed a classification technique which factors contextual information into the classification process using Markov models. It involves computing the a posteriori probabilities for all classes on a pixel-by-pixel basis. (This would be a by-product of running a minimum probability of error classifier, for example.) In training regions, transition probabilities (the probability that a pixel is class X given an adjacent pixel is class Y) are estimated for all classes, and used to sequentially update the a posteriori probabilities in small neighborhoods. The size of the neighborhood determines the spatial extent of the update process. After the probabilities have been updated, the class having the largest a posteriori probability of occurrence is assigned to each pixel. Improvements in classification accuracy of 5% and 7% in 3×3 and 5×5 windows, respectively are reported in Ref. 131.

4.5.3 Multi-Temporal Classification

Lennington (Ref. 126) has developed a method to estimate the proportion of small grains present in an image based on fitting values for the tasseled cap greenness (Ref. 123) over time to a vegetative growth model on a pixel-by-pixel basis. Pixel greenness is related to the amount of biomass (vegetation) present in the scene. Parameters of the vegetative growth profile such as the time when greenness peaks, the peak greenness value, and the time between inflection points about the peak are used as features for classification. Histograms of these features are then computed and decomposed into a sum of normal densities using the CLASSY algorithm discussed previously. A weather-driven model is used to predict acceptance ranges in feature space for small grains. Any distribution which falls into the range is classified as a small grains. Lennington claims that the technique is general enough to be used to obtain proportion estimates and classifications for other crops that are difficult to classify using single-look multi-spectral data. It is possible that the techniques may also be extendable to other sensors such as SARs.

4.5.4 Signature Extension/Normalization

In order to be able classify surface materials over a wide range of scene, sensor, and environmental conditions, robust representations (statistical, or otherwise) are needed. In attempting to use the training statistics computed in one image to classify another, signal variability can become a problem. For example, one image may be hazier than another, or images taken at different times may be different due to changes in the illumination, or in the biomass.

One approach to the problem has been through signal normalization (also called signature extension). Henderson (Ref. 132) uses a multiplicative and additive signal correction (MASC) algorithm to map signatures in one data set to those in another. He shows a significant reduction in classifier error rate using the technique. In order to use the MASC algorithm, clusters in one image must be matched to those in the other image. Henderson describes one method to accomplish this which involves clustering the images, ordering clusters according to their means, matching clusters on the basis of their means, and computing additive and multiplicative corrections for each band by linear regression. The problems with this approach, however, are that the matching of clusters must be supervised (thus reducing the level of automation possible in a system), and the correction may have to be spatially varying (to account for non-uniform haze, for example). An alternative to this approach is to perform a haze correction as described in Section 4.2.

4.5.5 Knowledge-Based Classification

A major step towards further automating the feature extraction process is the development of a classifier that does not have to be trained on an image-by-image basis. MASC eliminates the training phase (in subsequent classification), but substitutes the cluster matching step, which should be supervised.

An alternate approach under investigation at TASC is based on the use of relative image measures to characterize surface material classes in the form of rules. The approach, discussed in Appendix G, involves expressing the typical appearance of materials in terms of relative image measures. For example, if water is present in a scene, then the darkest

regions in the infrared are likely to be water. The approach allows expert knowledge of the domain to be used to develop rules for classification directly, thus eliminating the training phase and decreasing the amount of operator supervision required. The use of relative image measures also mitigates such effects as haze or sensor gain variations between scenes and/or images.

4.5.6 Summary

Table 4.5-1 summarizes representative techniques from each major category of image classification techniques discussed in this section. In general, pixel classifiers are the simplest to design and implement, but are not very efficient. Since region classifiers process groups of pixels at a time, depending on how expensive it is to group pixels into regions, region classification can be quite efficient (up to a 50% decrease in classification time has been reported in Ref. 130). Since region classifiers make use of information from neighboring pixels, classification accuracy can also be improved. Multi-temporal techniques increase the ability to discriminate between, and classify certain types of vegetation. Signature extension allows the spectral signatures of known material types in one image to be mapped to another. Since all of the above techniques require some degree of supervision (training or signature mapping), the degree to which the surface material classification process can be automated is limited. Knowledge-based techniques have the potential to further automate the process, however, additional work is required.

TABLE 4.5-1

SUMMARY OF IMAGE CLASSIFICATION TECHNIQUES

TECHNIQUE	DESCRIPTION	UTILITY
Pixel classification	Assign pixels to classes using minimum-distance or maximum a posteriori decision criteria	Simplest classification strategy
Region classification (ECHO)	Aggregates pixels by region-growing and assigns regions to classes by sample classification (e.g., hypothesis testing approach)	Improves classifier accuracy by taking information from neighboring pixels into account
Multi-temporal classification (CLASSY)	Compares pixel "greenness" in time to a vegetative growth model	Useful in discriminating between, and classifying vegetation types that are difficult to classify using "single-look" imagery
Signature extension (MASC)	Normalizes a data set by mapping clusters in one image to those in another image that has already been classified	Does not require training; supervision of cluster mapping process is recommended
Knowledge-based classification	Uses hueristic rules to characterize surface material classes; classifies image in a hierarchical fashion	Does not require training; domain knowledge may be used to develop classification rules directly

4.6 OBJECT IDENTIFICATION

Object identification follows surface material classification in the MS/MS feature extraction system and involves grouping regions into objects (possible DMA features), and identifying objects based on properties of the constituent regions. As part of Task 5 a review of three major computer vision systems was performed: Acronym (Ref. 140), the Kyoto University system (Ref. 110), and the University of

Massachusetts VISION system (Ref. 26). It was concluded that while many existing image processing, segmentation, and classification techniques are applicable to MS/MS feature extraction, few techniques appeared directly applicable to object identification.

In this chapter, a processing-flow for object identification using 2-d object models is developed, and candidate techniques are described. The 2-d approach assumes that the symbolic descriptions derived from 2-d projections of inherently 3-d scenes are sufficient for identification. Such an assumption is often acceptable in aerial imaging applications where the illumination is far from the scene, the view angle is relatively fixed over the field-of-view, and occlusion is not a significant factor (Ref. 141).

4.6.1 Review of Candidate Computer Vision Systems

Table 4.6-1 summarizes our assessment of the three vision systems mentioned above. Acronym (Ref. 140) was designed to be a generic model-based vision system. It divides the model-based vision process into four parts: modeling, prediction, description, and interpretation (Ref. 141). Objects are modeled by the volumes they occupy (generalized cones and cylinders) and by transformations between the local coordinate systems of these volumes (i.e., three-dimensional spatial relationships). Object classes are defined by constraining the actual parameter values (i.e., the lengths, widths, distances, etc) to be within certain ranges. Prediction involves projecting a model into the image plane to determine what features will be visible in the image and what their spatial relationships will be. This serves as a starting point for hypothesizing object-to-image feature matches. The image is described in terms of ellipses and ribbons (which are the two-dimensional

TABLE 4.6-1
ASSESSMENT OF CANDIDATE COMPUTER VISION SYSTEMS

SYSTEM	DESCRIPTION	ASSESSMENT
Acronym	Model-based vision system designed to be domain-independent; has been applied to aerial image interpretation and industrial parts inspection applications	Image-level processing and segmentation produces only edge-based symbolic descriptions which appear inadequate for feature extraction
Kyoto System	Prototypical system developed for analyzing color infrared (multispectral) photography; uses 2-d models to represent objects which may appear in a scene	Experimental system; although applicable to feature extraction it is somewhat limited in its capabilities
VISIONS	Vision system patterned after the Hearsay speech recognition system; contains long and short term memories organized in a hierarchical fashion	More refined than above system, particularly with respect to control strategies and richness of object descriptions; applicable to feature extraction

projections of generalized cones and cylinders). The Nevatia-Babu line finder was used to find the lines which made up ellipses and cylinders. Acronym interprets an image by finding subgraph isomorphisms between the prediction graph generated from the model and knowledge of the imaging geometry, and the picture-graph generated from the image. Acronym appears to be one of the (if not the only) complete vision system in use. Hughes has been attempting to expand the basic design, which used only edge information, for a submarine port monitoring application (Ref. 142).

A prototype system has been developed at Kyoto University for analyzing color infrared aerial photographs (Ref. 110). The system uses two-dimensional object models to represent the

kinds of objects which may appear in a scene. A two-dimensional approach is acceptable in aerial imaging applications since the viewing angle is relatively fixed over the field of view and occlusion is not a significant factor. In this system, cue regions (large homogeneous regions, shadows, etc) are used to trigger object recognizers. A region is classified by examining a number of simple properties. For example, if a region is large and spectrally homogeneous, is composed of vegetation, does not contain water, and is not a shadow-making region (i.e., is relatively low to the ground) then it is marked as a crop field. As object recognizers are triggered, they write their respective hypotheses into a short-term memory or blackboard. The blackboard also serves as a communication mechanism between object recognizers.

The blackboard concept was originally developed for speech recognition (Ref. 143). In the VISIONS (Ref. 26) system, the blackboard concept is further exploited with emphasis on developing control and interpretation strategies for image interpretation. VISIONS is like Hearsay in that knowledge is represented at a variety of levels. A short-term memory (STM) stores descriptions of the image in terms of vertices, segments, regions, surfaces, volumes, objects, and schemas. The long-term memory (LTM) contains the knowledge required to generate hypotheses at one level from information at other levels. Schemas are like contexts (e.g., a house scene) and are used to constrain the selection of object models stored in LTM. The interpretation process proceeds in a top-down fashion: models are projected into the image and matched against features derived from the image.

Of the above three systems, Acronym is considered to be least applicable. It uses only edge information that has

been derived from black-and-white imagery. Moreover, it requires that detailed 3-d models of all objects be defined. Object identification is performed by matching edge-based symbolic descriptions computed from the image to those predicted by projecting 3-d models onto the 2-d image. Both the Kyoto University vision system, and the University of Massachusetts VISIONS system make use of color imagery, and utilize region-based segmentation techniques to partition images into spectrally homogenous regions. VISIONS is more sophisticated, making use of hierarchical knowledge representations and control structures to characterize and identify objects.

4.6.2 2-D Object Recognition

This section addresses three key processes underlying object identification: connected region extraction, structural analysis, and grouping and identification. First, a surface material classification map is organized by grouping pixels with the same material type into connected regions. Next, selected attributes (geometrical, topological, and relational) are computed for grouping and identification. Based on these attributes, regions are organized into objects and objects are identified as instances of DMA features. In the section below, each of these processes is discussed in detail.

Connected Region Extraction - Connected region extraction is performed on a surface material classification map. The output is an image of labels, each corresponding to one connected region in the SMC map. The objective is to group contiguous pixels having the same SMC into a region with a unique label for spatial referencing by subsequent processes.

Prior to labeling the connected regions, it may be desirable to preprocess all or part of the SMC map to remove

small regions of little or no practical significance, to split large regions into smaller pieces, or to merge small regions into aggregates. For example, if large agricultural areas are sought after, small regions identified as crops and plowed fields can be removed by a shrink/expand operation. Shrink/expand operations can be easily performed by replacing the center pixel in a sliding window by the local minimum/maximum. Shrink/expand operations can also be used to fragment networks of interconnected pixels of similar SMC. An example is splitting urban areas (made up largely of concrete) joined by concrete roads into thin regions (roads) and large compact regions (urban areas). Shrinking obliterates all but the large blob-like regions, which are expanded back to their original size. Subtracting this from the original yields the small thin regions obliterated by the shrink operation. Expand/shrink operations are useful for merging regions that are relatively close to one another into aggregate regions (e.g., crops and plowed fields into agricultural regions).

Pavlidis (Ref. 133) discusses several methods for contour filling and region growing which may also be used to label connected regions in the image. One such technique labels the image recursively, starting at a seed pixel. It examines, in order, pixels above, below, to the left, and to the right of the seed pixel to see if a neighbor's value is the same as the seed's. If so, in a second array, the label of the seed is assigned to that of the neighboring pixel, and the algorithm calls itself at the neighbor's location. This process proceeds until all regions are labeled. The number of labels at the end is equal to the number of (in this case) four-connected regions in the image.

Structural Analysis - In some cases, DMA features will consist of single regions and may thus be classified on the basis of such properties as the region's area, perimeter, or compactness. Properties such as these are easily computed from labeled connected regions (Ref. 134). In order to recognize man-made and natural objects on the basis of their two-dimensional shape or silhouette, more complex descriptors such as hierarchical polygon decomposition (Ref. 135), curvature primal sketch (Ref. 136), and Fourier descriptors (Ref. 137) will be required.

Grouping and Identification - The output from the structural analyzer is a symbolic description of the image which enumerates the various attributes (size, shape, location, surface material type) for each region. Prior to object classification it may also be necessary to group regions into objects. This can be done on the basis of relative proximity (Ref. 138), and other local properties. For example, if two regions are relatively close to one another and have similar orientations (this can be determined by analyzing histograms of the orientation of, and distance between regions), they are grouped into an object. This kind of bottom-up grouping underlies Marr's theory of texture vision (Ref. 139). An alternate approach to grouping involves aggregating regions into objects based on prior knowledge concerning the kinds of regions likely to form the object.

Once the symbolic description has been organized, rules may be applied to each object to determine its identify based on the properties of its constituent regions. For example, in Acronym objects are modeled by the volumes they occupy (generalized cones and cylinders) and by transformations between the local coordinate systems of these volumes (i.e., 3-d

spatial relationships). Object classes are defined by constraining the actual parameter values (i.e., the lengths, widths, distances, etc) to be within certain ranges. On the other hand, in the Kyoto system 2-d object models are used to represent the kinds of objects which may appear in a scene. If, for example, a region is large and spectrally homogeneous, is composed of vegetation, does not contain water, and is not a shadow-making region (i.e., is relatively low to the ground) then it is identified as a crop field.

4.6.3 Summary

In this section, a review of three candidate computer vision systems for MS/MS feature extraction was performed. A process for identifying objects was described which involves grouping pixels with the same surface material type into connected regions, computing 2-d attributes such as the size and shape of, and relations between regions in the image, aggregating regions into candidate objects based on relative proximity, collinearity, and composition, and identifying region(s) as instances of predefined objects by expressing the typical (or expected) appearance of the object in terms of the above attributes in the form of rules. Such an approach has been shown to be appropriate in aerial imaging applications where the illumination is far from the scene, the view angle is relatively fixed over the field of view, and occlusion is not a significant factor, and is thus appropriate for use in interpreting imagery from space-based sensor systems such as Landsat.

4.7 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

4.7.1 Summary

This chapter has reviewed representative multi-spectral and synthetic aperture radar (SAR) sensors, and assessed the use of image processing, segmentation, classification, and object recognition techniques in exploiting the data provided by these sensors for feature extraction.

4.7.2 Conclusions

Of the four sensor systems reviewed: LANDSAT, SPOT, SEASAT SAR, and the Shuttle Imaging Radar (SIR), the LANDSAT MSS is the only multi-spectral sensor system to have achieved operational status; the other sensors were largely developmental in nature. However, a spatial resolution of 56 × 79 m in the visible and reflective IR limits the use of the MSS in feature extraction applications. The TM has both superior resolution (30 m visible and reflective IR), and spectral resolution (7 bands, including thermal IR) over the MSS. The added bands gives the TM improved ability to discriminate geologic resources, types of vegetation, and land use. However, its resolution still limits its use to compiling and updating large scale maps only. The planned French SPOT satellite will have better spatial resolution (20 m multi-spectral and 10 m panchromatic), but will have fewer spectral bands (only 3).

The SAR sensors (SEASAT SAR, SIR-series) reviewed are all considered to be experimental in nature. All are L-band radars (23.5 cm) with a spatial resolution of about 25 m (four looks averaged). Both the SEASAT SAR and SIR-A were uncalibrated devices. In exploiting this type of SAR imagery one must be aware of possible limitations in dynamic range of the

data (typically 4 bits), and must be prepared to deal with a considerable amount of "speckle". Although comparable in spatial resolution to the TM, cartographic accuracies for these sensors are lower, at least ±50 m (and possibly less, depending on how careful the user is in selecting control point pairs).

Among the image classification techniques discussed in this section, pixel classifiers are the simplest to design and implement, but are not very efficient. Since region classifiers process groups of pixels at a time, depending on how expensive it is to group pixels into regions, region classification can be quite efficient (up to a 50% decrease in classification time has been reported in Ref. 130). Since region classifiers make use of information from neighboring pixels, classification accuracy can also be improved. Multi-temporal technqiues increase the ability to discriminate between, and classify certain types of vegetation. Signature extension allows the spectral signatures of known material types in one image to be mapped to another. Since all of the above technqiues require some degree of supervision (training or signature mapping), the degree to which the surface material classification process can be automated is limited. Knowledge-based techniques have the potential to further automate the process, however, additional work is required.

Of all MS/MS technology areas, image processing appeared to be the most mature. Past work in remote sensing has provided many techniques of potential use in feature extraction.

In some cases, techniques that were originally developed for optical imagery are directly applicable to multispectral and multi-source imagery. For example, geometric transform techniques developed for optical imagery are useful

for registering SAR and multi-spectral data sets as well. On the ohter hand, while black-and-white (single image) enhancement techniques can be applied on a band-by-band basis, new techniques which exploit correlations between bands (for thermal band sharpening) and between sensors (for using coregistered optical imagery to smooth SAR) appear promising.

Finally, new image transformations based on the tasseled cap and canonical variates approach which provide physically-significant information (e.g., vegetative cover, soil moisture) can be expected to be of considerable utility to the image anlaysis in manual interpretation as well as in surface material classification.

Several computer vision systems were assessed as candidate automatic MS/MS feature extraction systems. Systems developed at Kyoto University (Ref. 110) and the University of Massachusetts (Ref. 26), which use 2-d models to represent objects of interest in a scene, appeared applicable. Although additional developments in this area will be necessary before an automatic system can be developed, the 2-d approach did appear promising for feature extraction. Such an approach has been shown to be applicable in aerial imaging application where the illumination is far from the scene, the view angle is relatively fixed over the field of view, and occlusion is not a significant factor.

4.7.3 Directions for Further Research

While image pre-processing is considered to be a fairly mature technology area, further work in several areas is recommended. First, our assessment of image restoration and enhancement techniques revealed that while many single-band (monoscopic) techniques exist, few make explicit use of more than one band

or sensor. Initial results presented in Appendix F demonstrated the utility of multi-band/sensor techniques for spatial enhancement. It is recommended that additional work be performed to quantify the performance of multi-band thermal band sharpening and SAR smoothing techniques, and to investigate other applications of the technique. (One such use for detecting and restoring data drop-outs was suggested in the report.) The use of tasseled cap transforms and canonical variates to extract information such as vegetative cover, wetness, and concreteness from an image, should also be pursued. In particular, transforms for other physically-significant properties should be derived.

Alternate image classification strategies (e.g., knowledge-based techniques as described in Appendix G) need to be more fully developed and tested. An operational assessment of different image classification strategies (with ground truth) should be conducted to determine the merits of heuristic versus statistical techniques (i.e., to what extent can heuristic rules increase the level of automation possible in the classification process), to determine to what extent region-based classification is superior to pixel-based classification (e.g., in terms of error rate, and processing time), and to determine to what extent prior information (e.g., context) improves classifier accuracy. The assessment should be performed using a variety of scenes (agricultural, residential, and urban), acquired at different times (time of day and season), and under a variety of scene/sensor conditions (haze, sensor noise levels).

Finally, it is suggested that a testbed be assembled for assessing MS/MS feature extraction techniques. (The RWPF-upgrade would be a candidate target system.) The objectives of the testbed would be three-fold: to allow experimentation

with diverse imagery sources to determine what kinds of information can be readily extracted from what types of imagery under what conditions, to allow prototypical feature extraction systems (i.e., special-purpose vision systems) to be developed and tested, and to provide an environment for DMA to transition new these new feature extraction technologies into production systems.

5. CONCEPT OF OPERATION FOR A MULTI-SPECTRAL/ MULTI-SOURCE (MS/MS) FEATURE EXTRACTION SYSTEM

In this chapter a concept of operation for an interactive, semi-automated MS/MS feature extraction system based on current image processing, analysis and computing technologies is formulated. Under Task 4 a concept of operation was developed which was based on the use of black-and-white feature extraction techniques identified and assessed in Task 3. The latter concept of operation assumed that no reliable, fully-automated approach to feature extraction would be feasible before 1985. The approach taken to developing the concept of operation, therefore, employed a semi-automatic implementation which made use of available feature extraction techniques to improve DMA's current feature extraction process. Improvements would result mainly through the use of

- Interactive image enhancement to permit an image analyst to see, measure and classify features of interest more easily
- Semi-automatic screening to determine whether features of interest are likely to be present in selected areas
- Partial feature detection, which would cue operators to areas requiring further analysis, and reduce the likelihood of undetected features
- Highly local, directed scene analysis coupled with photogrammetric models to locate, delineate and measure detected features
- Highly local, directed pattern recognition or feature description expert systems to classify features.

The resultant concept of operation for a semi-automated feature extraction system was formulated as follows:

- A generic, implementation-independent concept of operation for feature extraction was developed to facilitate the identification of functions and activities which could potentially benefit from automation in general, and machine perception technology in particular

- Techniques and technology assessed in Task 3 were related to selected activities within the generic operational concept, and alternative approaches for automating them were described

- A feasibility and cost/benefit analysis was performed for each proposed approach and the most promising methods for automating the selected feature extraction activities were identified

- A refined concept of operation was formulated based on the results of the latter analysis, and a candidate system architecture identifying possible hardware/software components for implementing the concept of operation was developed.

In Task 5 it was concluded that the process of classifying surface materials in MS/MS imagery could be automated to some degree by 1985. It was also pointed out that once an image had been partitioned into surface material classes, other attributes such as 2-d shape and size could be extracted and used to support feature identification. Hence, the approach taken to developing a concept of operation for a MS/MS feature extraction system was:

- To continue to exploit interactive image enhancement techniques to permit the image analyst to more easily see, measure and classify features of interest in MS/MS imagery

- To establish as a primary goal in semi-automatic feature extraction scenarios, the classification of surface materials present in the scene
- To use the resultant surface material classification map to drive the object recognition process (i.e., to group regions into potential features or objects based on a priori knowledge concerning the physical composition of those features) and to support an image analyst and/or expert system in inferring the identify of a feature based on attributes derived from it.

The concept of operation described in this chapter describes in detail the activities performed within the following three functional areas identified in Chapter 4:

- Pre-Processing
- Surface Material Classification
- Object Recognition.

In each, activity-flows are constructed which describe how the various techniques are used to support the feature extraction process. Alternate process-flows and processing options are also considered whenever it is appropriate to do so.

The feature extraction process is organized into three major activities: pre-processing, surface material classification, and object recognition. Conceptually, all data processing functions within the system communicate via the system database as shown in Fig. 5-1. It is assumed that source material and user requirements are initially examined during the pre-processing activity and disseminated throughout the system via the system database. In this and other activities-flows depicted in this chapter, solid lines denote data-flows, and dotted lines denote control-flows within the system.

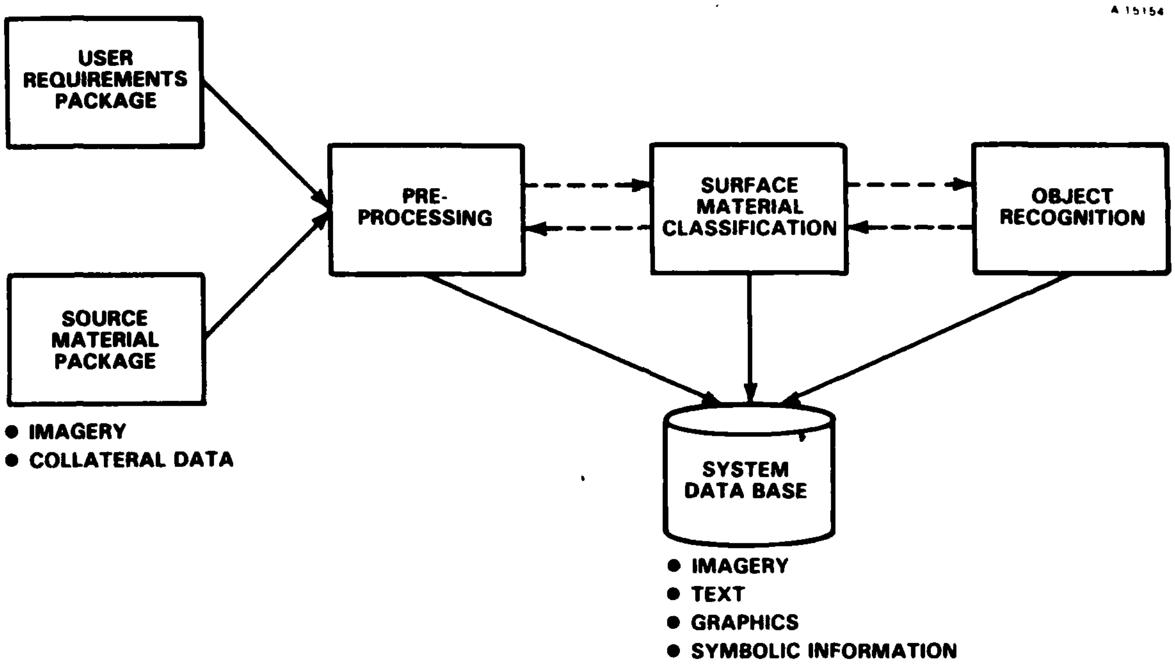


Figure 5-1 Organization of Functional Areas within the MS/MS Feature Extraction System

The remainder of this chapter is organized as follows. Section 5.1 addresses how image processing (i.e., registration, restoration, and rectification) techniques may be used to exploit MS/MS imagery interactively as well as to prepare MS/MS imagery for subsequent machine processing. In Section 5.2, several alternate process-flows for surface material classification are presented. Issues such as efficiency, accuracy and robustness are assessed for each. The use of MS/MS imagery to support the identification of cultural and terrain features is outlined in Section 5.3. A brief summary of data processing requirements in the above three areas is contained in Section 5.4. Based on results from Sections 5.1, 5.2, and 5.3, a candidate architecture for an all-digital MS/MS feature extraction work station is provided in Section 5.5. Finally, Section 5.6 summarizes our conclusions and provides recommendations for future experimentation and research in this area.

5.1 PRE-PROCESSING

Pre-processing serves two major purposes in the MS/MS feature extraction system: to prepare multiple and diverse imagery sources for subsequent semi-automatic interpretation (i.e., surface material classification and object recognition), and to provide a set of highly interactive display and image processing capabilities to support manual (computer-assisted) interpretation. Pre-processing functions include functions for registering, restoring, enhancing, and transforming MS/MS data. The goal is to provide imagery which is normalized in the sense of being registered to a common coordinate system, having sensor and atmospheric artifacts removed (e.g., destriped and haze corrected), and having the same effective ground resolution, perceivable contrast, and noise level.

Pre-processing involves the following major activities:

- Source Assessment: Assess the quality of the imagery (contrast, noise level), and the quality of the coverage (percent cloud-cover, haze) to determine to what extent the source imagery is usable

- Registration: Bring imagery acquired at different times and by different sensors into pixel-by-pixel alignment; register to map or chart

- Restoration: Detect and replace imagery data lost in transmission, and correct for degradations in spatial sharpness and contrast due to atmospheric and sensor effects

- Enhancement: Accentuate tonal, textural, and spectral properties for human interpretation

- Image Transformation: Convert raw spectral/temporal data into alternate forms (principle components, tasseled cap transforms) to support image classification.

Outputs from source assessment are sent to planning activities for surface material classification and object recognition to determine if the available source will support semi-automatic processing. If so, appropriate steps are taken to prepare the imagery for subsequent machine interpretation. Otherwise, manual interpretation may be performed. Registration, restoration, enhancement, and image transformation functions are performed interactively. The typical process-flow is shown in Fig. 5.1-1; however, in practice the order in which the various functions are executed may be different. In all activities, rapid access to image and collateral databases and highly interactive image-oriented workstations are required.

5.1.1 Source Assessment

Source assessment represents the initial planning activity in the MS/MS feature extraction system since its function is to first determine whether or not the available source material is sufficient to satisfy the specified interpretation requirements (e.g., spectral/spatial resolution, and coverage), and second, to determine to what extent semi-automated techniques may be employed. As mentioned earlier, there can be expected to be a considerable amount of communication between planning activities in pre-processing, surface material classification, and object recognition. Object recognition processing requirements shall affect what surface material classification strategy to use. Similarly, the requirements of surface material classification will affect the type of pre-processing needed. Manual interpretation scenarios can be less structured in the sense that the user should be able to decide on the fly what enhancement technique, for example, should be used on a particular image.

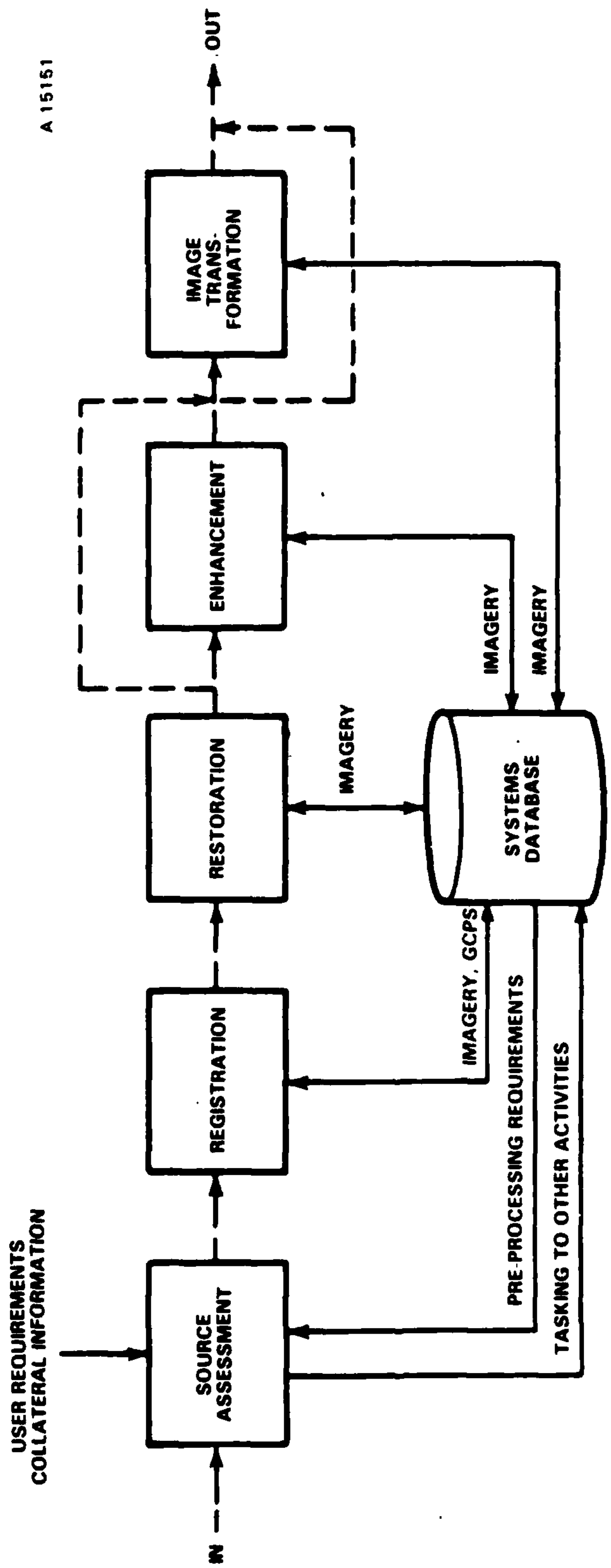


Figure 5.1-1 Pre-Processing Activity-Flow

5.1.2 Registration

The user may select from two options in registration:

- Relative Registration: Register two or more images of the same scene to one another. This is sufficient for subsequent data extraction activities, and allows the set of raw MS/MS image measurements to be spatially accessed as vectors.
- Absolute Registration: Register one or more images to a geodetic coordinate system in order to use extracted data for map generation and revision. Absolute registration thus involves the added step of registering MS/MS data to collateral source material such as a map or chart.

Both options require the selection of ground control points (GCPs) in order to perform registration. For absolute registration, these GCPs must include points whose locations in the desired coordinate system are accurately known. Assuming a global polynomial (rubber sheeting) transformation is used for registration, the minimum number of GCPs needed for a given order can be determined (e.g., at least three GCPs are needed to compute a first order remapping polynomial). What remains to be done then is the selection of corresponding GCPs in each image.

One semi-automatic GCP selection strategy (Ref. 159) involves the following four steps:

(1) Coarsely register the imagery

(2) "Whiten" the imagery. (This involves applying a filter which sets the magnitude of the Fourier transform of the image equal to a constant without altering the phase of the image.)

(3) Perform short-space correlations over a regular grid. (Compute the offset of the "best match" relative to grid points and its confidence. Measures for confidence may be obtained as a function of the normalized correlation coefficient.)

(4) Select GCPs at a specified level of significance.

One example of this selection process is shown in Fig. 5.1-2. Selected GCPs for registering Landsat TM (Fig. 5.1-2a) to Seasat SAR (Fig. 5.1-2b) are shown in Fig. 5.1-2c. These GCPs are used to compute the remapping polynomial. The registered Landsat TM/Seasat-SAR image is shown in Fig. 5.1-2d.

5.1.3 Restoration

Typically, the type of restoration to be performed upon MS/MS data will depend both on the scene and the sensor(s) used. For example, haze correction is scene-dependent, while destriping and spatial sharpening are sensor-dependent. The actual techniques selected will depend on the nature of the degradation, and on the types and combinations of sensors used.

Among the available processing options possible are:

- Haze Correction: To correct for Rayleigh (short wavelength) scattering, pick an area in the image that has a high degree of correlation between bands, perform a regression analysis between bands to determine the offset (additive haze) component, and subtract.
- Destriping: To correct for non-uniform gain and offsets between detector arrays, apply a non-linear histogram remapping technique.

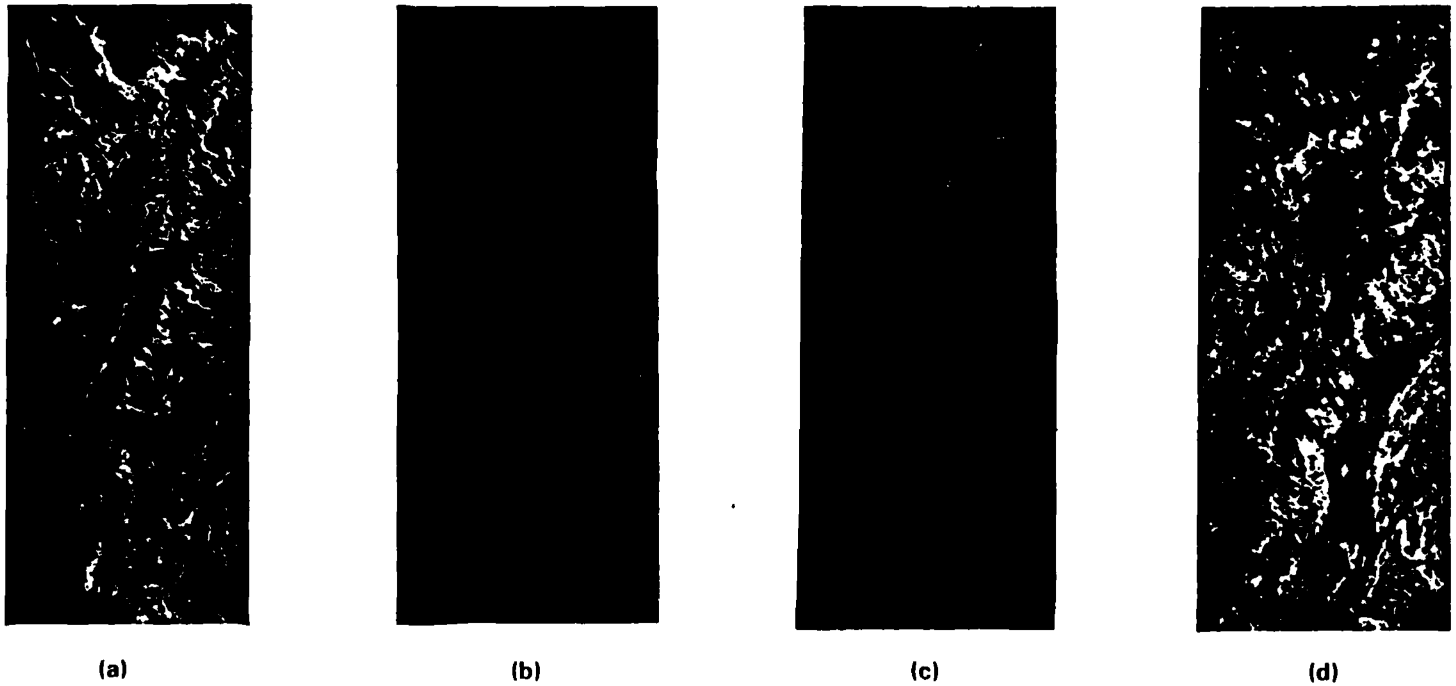


Figure 5.1-2 Registration of Multiple Imagery Sources: (a) Thematic Mapper, (b) Seasat-SAR, (c) Statistically Significant Control Points, (d) Registered TM/SAR Image

- Spatial Sharpening: To restore spatial sharpness lost due to finite aperture (MTF) effects, use a short-space high-frequency emphasis filter (blind restoration), or if transfer function and noise models are known, compute and apply a Weiner (minimum mean-squared error) filter; if multiple spectral bands/sensors are available, multi-band/sensor enhancement techniques may be applied (e.g., Landsat TM thermal band sharpening and SAR smoothing for speckle reduction as described in Appendix F).

Depending on the peculiarities of the sensor, other types of restoration techniques (e.g., pixel or line dropout detection/filling algorithms) may be appropriate to use as well.

5.1.4 Enhancement

Enhancement functions are used to adjust the overall tonal (color) balance of an image for display, and to increase the local perceivable information to aid in manual feature discrimination. Several techniques developed to enhance single-band (black-and-white) imagery were also useful for MS/MS imagery:

- Global Histogram Equalization/Contrast Stretching: Used to alter the overall tonal (color) balance of an image for display. Global histogram analysis and intensity remapping is performed on a band-by-band basis. Intensities may be remapped such that the output intensity histogram matches a reference histogram (histogram matching is often used to normalize images before they are compared), or such that the output dynamic range (contrast) is maximized. The latter is often done to increase the contrast of visible imagery (e.g., TM bands 1-3) degraded by atmospheric haze.

- Local Histogram Equalization/Contrast Stretching: Performed on a band-by-band basis to increase the local perceivable information content. Intensity remapping is performed on a local basis, either to equalize the local histogram, or to stretch the local contrast. Although such processing improves texture discrimination, certain types of artifacts may be introduced (e.g., halos around object boundaries). A local histogram equalization of Landsat MSS bands 4, 5, and 6 is shown in Fig. 5.1-3 (Ref. 53). The image was acquired over Saudi Arabia. This enhancement allows ships at sea, a smoke plume, and fine structure (roads) in urban areas not visible in the original data to be seen clearly in the enhanced imagery.

- Short-Space Image Sharpening: Frequency domain filters may be used to emphasize high spatial frequencies to enhance edge structure and texture on a band-by-band basis. Short space filtering may also be used to enhance man-made objects such as road networks, agricultural fields and urban areas.

Since the above techniques are performed band-by-band, the overall tonal balance may be affected. To reduce color distortion, instead of enhancing multi-spectral imagery directly, selected bands can be first transformed into an alternate color-space (e.g., into intensity, hue, and saturation components), and only the intensity component enhanced.

Since typically more than three spectral bands/sensors are available, the user must assign one band to each primary color for display. Several color assignments are common. Two are shown in Fig. 5.1-4. Displaying TM bands 1-3 in blue, green, and red provide an almost true-color representation of the scene (Fig. 5.1-4a). By displaying bands 2, 4, and 7 in blue, green, and red (Fig. 5.1-4b) vegetation is easily recognized as green areas in the image (since vegetation produces strong

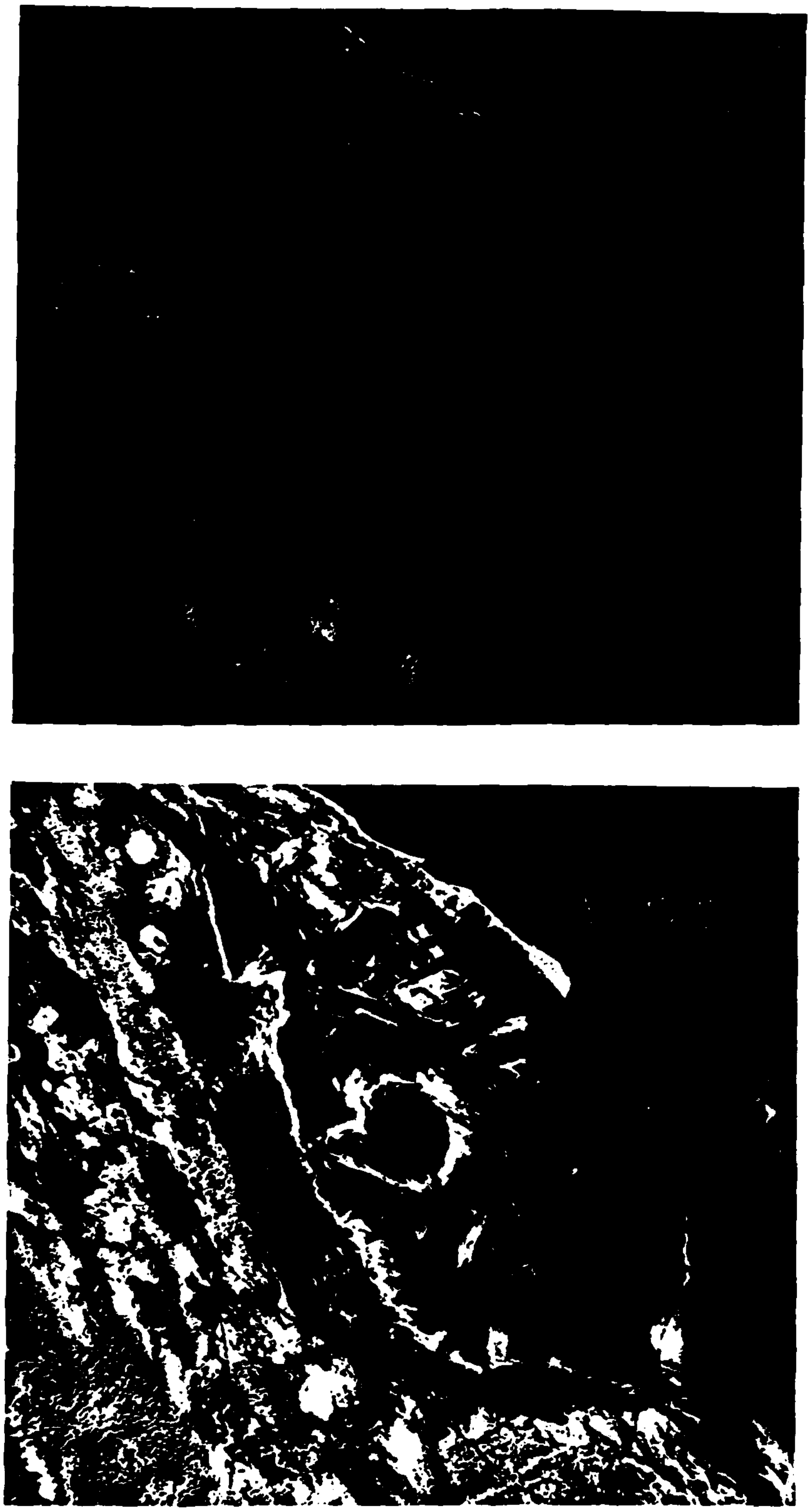


Figure 5.1-3 Landsat MSS Image (Bands 4,5, and 6) Enhanced Using Local Area Histogram Equalization

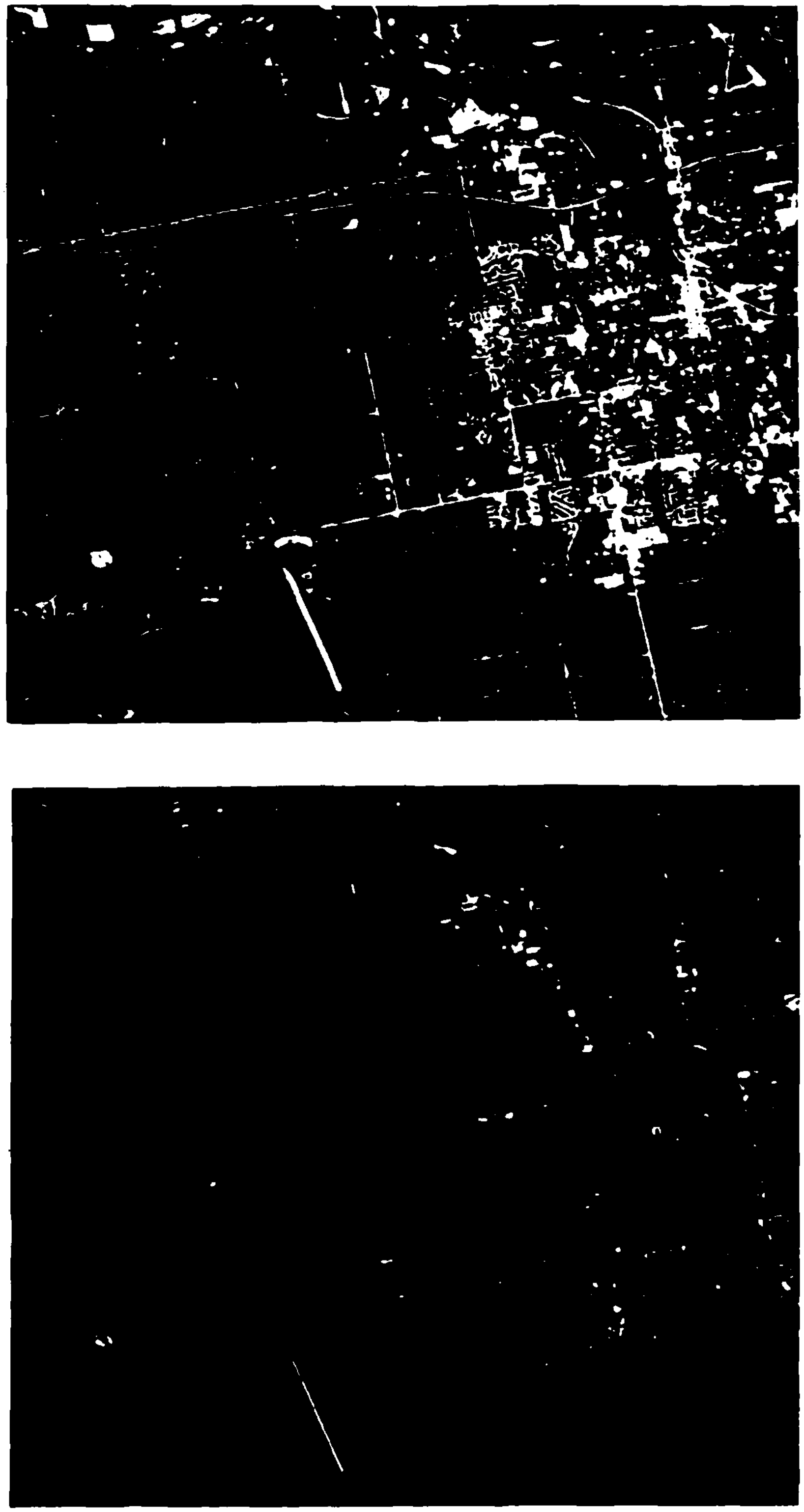


Figure 5.1-4 Landsat Image Acquired Over Lawrence, Kansas: Bands 1, 2, and 3 (Top) and 2, 4, and 7 (Bottom)

returns in band 4), and soil-like materials as red areas (due to the high reflectance of soil-like materials in the infrared), for example. Multiple display windows containing different band-color combinations would be highly desirable in a work-station environment.

5.1.5 Image Transformation

While enhancement is performed to improve the imagery analyst's ability to detect, identify, and delineate features, image transformation is performed to condition MS/MS data for subsequent machine processing. Two basic kinds of image transformations may be selected by the user:

- Linear Transformations: Examples include principal components, tasseled cap transformations, and canonical variates
- Nonlinear Transformations: Ratios, and nonlinear transformations of ratio images.

Prior to applying a statistical (decision theoretic) classifier, a principal components analysis and transformation is often performed to reduce the dimensionality of the decision space to reduce the cost of classification, and to orthogonalize the decision space to reduce the complexity of the classifier. To extract physically-meaningful image measures for other kinds of classifiers, a tasseled cap transformation or canonical variates analysis can be performed. In areas of considerable terrain relief, and thus shadows, ratioing can be used to reduce illumination variations for surface material classification. Nonlinear transformation of ratio images is typically performed to compress the result to a convenient dynamic range.

5.2 SURFACE MATERIAL CLASSIFICATION

Surface material classification is the process of inferring the physical composition of surface materials in an image from MS/MS measurements computed during pre-processing. It involves (optionally) segmenting imagery into regions having similar spectral properties prior to classification, and identifying pixels/regions as instances of pre-defined surface material classes. Measurements may be derived from multiple spectral bands/sensors or from data collected at different times. Statistical and heuristic inferencing techniques may be employed, and decisions made based either on prior knowledge concerning the typical/expected appearance of surface materials in the imagery or on training statistics.

Surface material classification involves the following major activities depicted in Fig. 5.2-1:

- Planning: Assess exploitation requirements (i.e., what materials would one like to extract at what resolution) against the available imagery (e.g., Landsat-TM), available techniques (what types of classifiers are available), and collateral information (e.g., signature libraries, training statistics, expert knowledge) to determine what classification technique/strategy to use.

- Classifier Development: Develop a classifier for the current scene in one of several ways: training on regions of known surface material type, mapping multi-spectral signatures derived in one image to the current image, predicting the signatures of selected surface material types.

- Classifier Evaluation: Test the classifier in areas where the ground truth is known, or alternatively, estimate the performance of the classifier by assuming

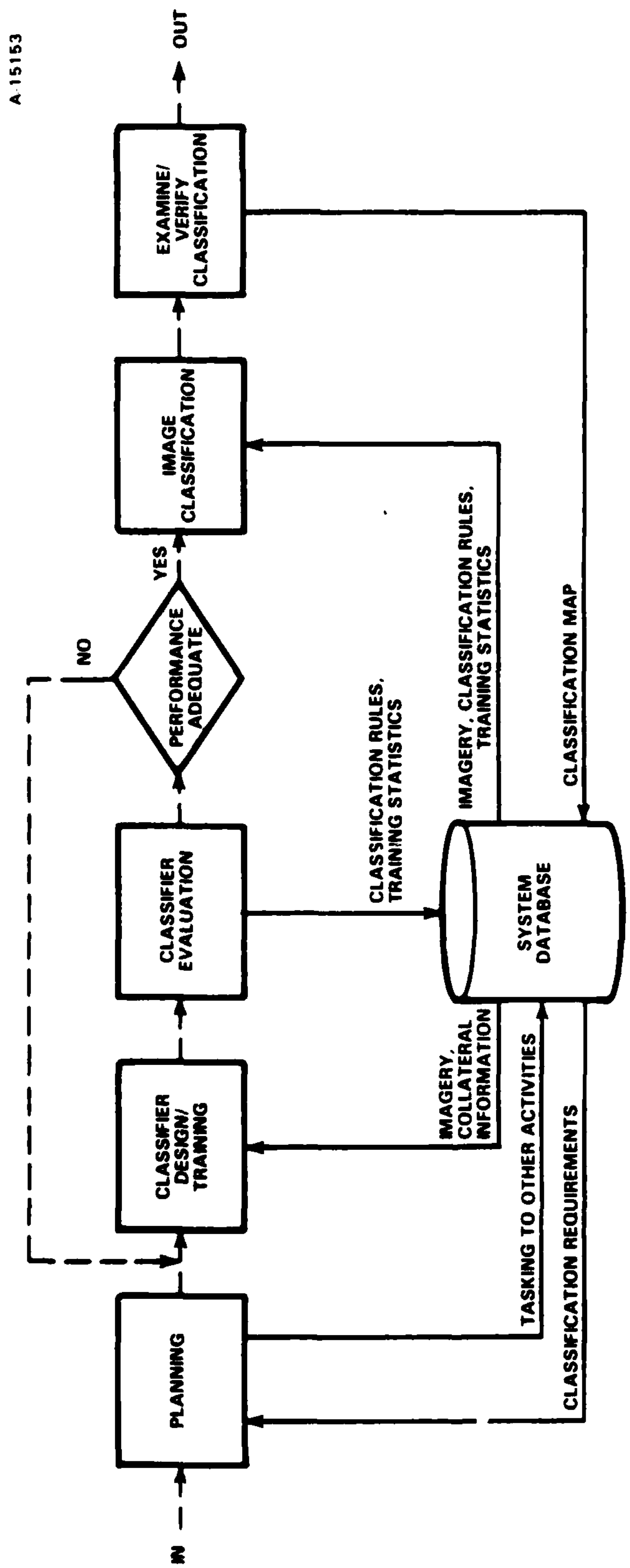


Figure 5.2-1 Surface Material Classification Activity-Flow

a statistical model for the class-conditional distributions, and compute a probability of misclassification.

- Perform Classification: Having achieved a desired performance on a test data set, apply the classifier to the full image; monitor classifier performance.

- Examine/Verify Classification: Display classification as a thematic map on a color display; verify classification using collateral data sources.

As shown in Fig. 5.2-1, the inputs to planning include collateral information (sensor parameters, types of materials expected to occur in the scene based on past observations, old maps (optimal) for use during the classification process), and exploitation requirements (what kinds of surface materials would one like to extract at what resolution). Classifier design/training involves selecting a classification strategy, and creating a classifier (instantiating a rule set, or defining decision regions based on training statistics). The instantiated rule set/decision functions are subsequently tested over an area of ground truth (or error bounds are estimated if ground truth is not available), and applied to the full the image. The resulting classification or thematic map is displayed. The user performs final verification and editing if required. Each of these activities is detailed below.

5.2.1 Planning

Planning involves an examination of the imagery relative to the requirements of the end-user (or product database) to determine if the source is adequate and if the knowledge base contains sufficient information to support the required exploitation. Initially, collateral information such as the date and time that the imagery was acquired, the quality of

the imagery, and the location of the scene can be used to determine if the information currently in the knowledge base is directly usable, or if training needs to be performed. If training needs to be performed, the availability of maps and charts must be determined.

5.2.2 Classifier Development

Once the characteristics of the data, the exploitation requirements, and the available knowledge have been assessed, a classification technique/strategy is selected. Depending on the type of classifier used, a variety of alternate process-flows are possible:

(1) Delineate regions which are representative of the surface material types of interest in the image. Compute training statistics or probability distributions (depending on the type of classifier to be used) over each region

(2) Segment/cluster the image to locate spectrally homogeneous regions. Associate (by hand) selected regions with instances of surface material types, and compute class statistics/probability distributions

(3) Segment/cluster the image to locate spectrally homogeneous regions. Associate clusters with instances of surface material types whose statistics are known in another image, and map the statistics into the current image.

(4) Given the kinds of materials likely to appear in the scene and knowledge with regard to the expected appearance of these materials in the scene, instantiate a rule set for the current scene.

Options (1) through (4) are arranged in order of increasing degree of automation. Options (1) and (2) both require ground-truth, but option (2) frees the user from having to select homogeneous regions that are representative of the surface material type. Option (3) requires that the man/machine be able to map the clusters for one image to those of another as is required, for example, in signature extension (Ref. 132). Option (4) is the most attractive since it requires neither training nor having to relate regions/surface material types between images.

5.2.3 Classifier Evaluation

If a subset of the data with known surface material composition can be isolated, the classifier may be tested and the performance evaluated empirically (e.g., using a confusion matrix). When this is not possible or practical, bounds on the classifier error rate must be estimated. In either case, if classifier performance is unacceptable, retraining should be performed.

A confusion matrix is obtained by classifying the training data set. High confusion rates between classes may be indicative of poor spectral separation, or a poor choice of training regions (i.e., the regions may not have been selected carefully enough). Several courses of action are possible: either retrain by selecting regions that are more precise or representative, or transform the data. For example, by performing a principal components analysis and transformation of the imagery, the dimensionality of the data set can be reduced significantly. Landsat TM data can typically be reduced from seven spectral bands to three or four principal components. By reducing the dimensionality, the error rate can often be

reduced. (Stated another way, an increase in the dimensionality can result in an increase in the error rate.)

5.2.4 Perform Classification

Classification may be performed either by pixel or by region. Regions may be formed prior to classification by region-growing or clustering. The advantage of first segmenting the image into regions is that since the number of regions is typically far less than the number of pixels, region-based classification will generally be much faster than pixel-based classification. To illustrate the tradeoffs, the following three statistical classification strategies are compared:

(1) Minimum-Distance Classification: Rank order the distances to each class mean in feature space, and assign the class that is closest to the value of the pixel or, for region classification, the mean of the region.

(2) Maximum-Likelihood Classification: Select the class having the largest a posteriori probability. Such an approach is compatable with algorithms which iteratively update a posteriori probabilities using information from neighboring pixels or semantic constraints.

(3) Hypothesis Testing (Non-parametric): A non-parametric method, such as the Kolmogorov-Smirnov test is used to compare sets of sample distributions. One set is of known origin (obtained through training), the other of unknown origin (computed over a region in the image). Classification is performed at a specified level of significance.

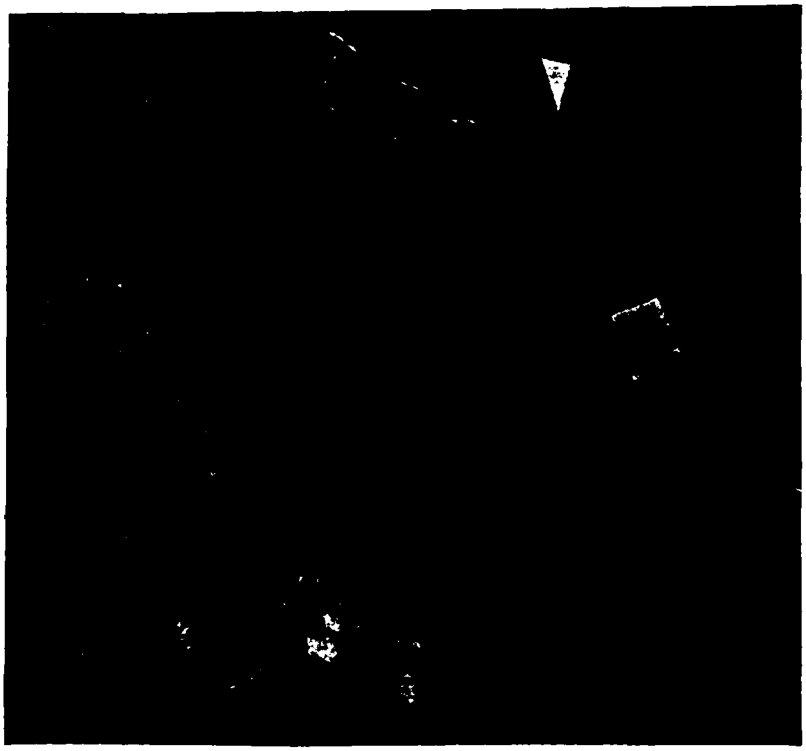
Method (1) may be used to classify pixels or regions. In general, the computational complexity grows with the number of classifications (i.e., the number of pixels or regions), as

well as the number of classes. While classification is performed on a pixel basis in (2), local (region) information may be used to update a posteriori probabilities. Depending on the application, a reduction in the error rate can justify the added computational cost (which can be quite high). A nonparametric method such as (3) compares sample distributions, which are computed over regions. The advantage to (3) is that by specifying a level of significance for the test, the performance of the classifier can be controlled. A region will be classified only if distributions are relatively similar, thus rejecting outliers. Also, if the two or more hypotheses are accepted (an indication that the reference classes are not well-separated), the null hypothesis, "no classification possible", can be selected to reduce the error rate further. An application of the technique to texture classification is described in Ref. 160.

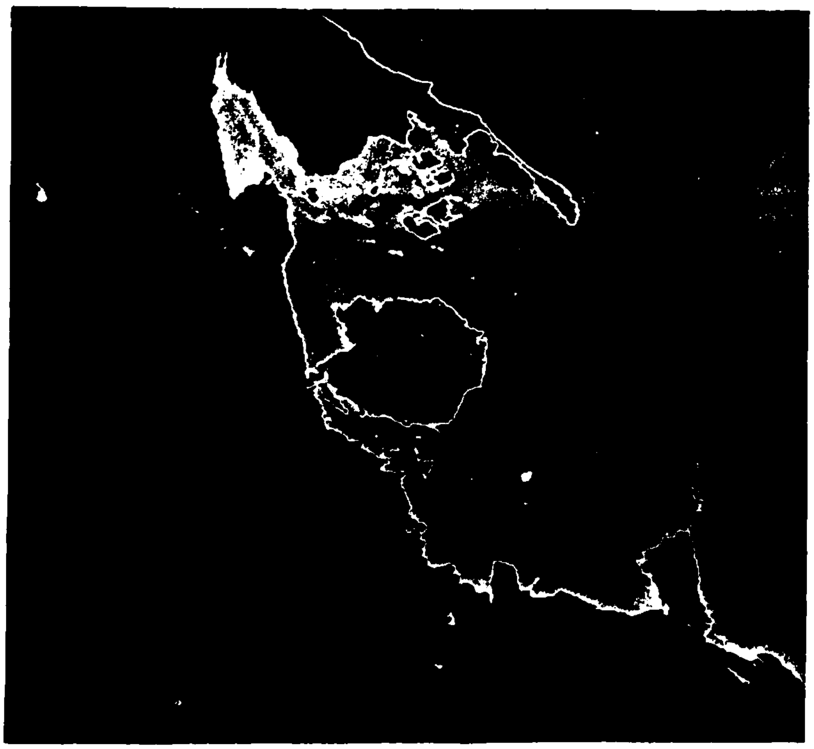
5.2.5 Examine/Verify Classification

After, or during classification, the results of the classification may be displayed in the form of a thematic map. Performance information (e.g., classification confidence) may also be provided. For example, a measure of the similarity between a pixel/region and the chosen class (e.g., minimum distance, or a posteriori probability) can be displayed. In an image of minimum-distances, dark (bright) regions indicate high (low) classification confidence, i.e., a pixel was relatively close to (far from) the selected class mean.

The example presented in Fig. 5.2-2 illustrates the general classification process described in the preceding sub-sections. In Fig. 5.2-2a the regions used to train the classifier are shown. Four classes were selected: vegetation (green), desert (yellow), culture (red) and water (blue). The



(a) Training Regions



(b) Minimum Distance Classification

Figure 5.2-2 Landsat MSS Classification Example

(c) Classification confidence

(d) Results (b) and (c) Combined

Figure 5.2-2 Landsat MSS Classification Example (Continued)

resulting classification and confidence images are shown in Figs. 5.2-2b and 5.2-2c. These results are combined in Fig. 5.2-2d with class and confidence encoded as color and saturation, respectively. In Fig. 5.2-2d saturated colors denote high classification confidence.

5.2.6 Comparison of Classification Strategies

Depending on processing requirements (level-of-detail, resolution), the type and quality of the available imagery, the level-of-automation desired, the availability of ground truth, and the computational resources available, several classification strategies are possible:

(1) Supervised Training Followed by Pixel Classification: User delineates homogeneous regions using collateral data (e.g., a map). Training statistics are computed. At each pixel, a minimum distance classifier assigns pixels to the nearest class.

(2) Signature Extension/Pixel Classification: An image containing similar material types has been classified, and the class-conditional distributions/statistics for each class (cluster) have been rank-ordered in feature space. The image that is to be classified is clustered using an iterative cluster-splitting approach. (The optimal number of clusters in obtained by splitting clusters until a measure of cluster quality is maximized.) Clusters in the two data sets are matched in a supervised fashion by comparing respective means and covariances. The data set is normalized using the MASC algorithm. Pixel classification is subsequently performed.

(3) Region Growing/Classification: Region-growing is performed to extract spectrally-homogeneous regions. The process is supervised so as to allow the user to control spectral homogeneity; i.e., how

much spectral intensities are allowed to vary within a region. The resultant regions are then classified on the basis of their sample distributions.

(4) Rule-Based Pixel Classification: Heuristic knowledge concerning energy/matter interactions, atmospheric effects, and sensor models are used to develop classification rules. The rules are applied on a pixel-by-pixel basis.

Table 5.2-1 summarizes our assessment of the above four classification strategies.

In general, the computational cost required in each varies strongly with the number of classes and dimensionality of the data set. The latter will depend on the number of spectral bands/sensors used, and on the kind of pre-processing performed (e.g., the use of principal components or tasseled cap transforms to reduce the dimensionality of the data set). Performance is largely dependent on the statistical separability of the various classes; the within-class scatter (variance) should be low and the between-class scatter (distance between class means) should be high for best performance.

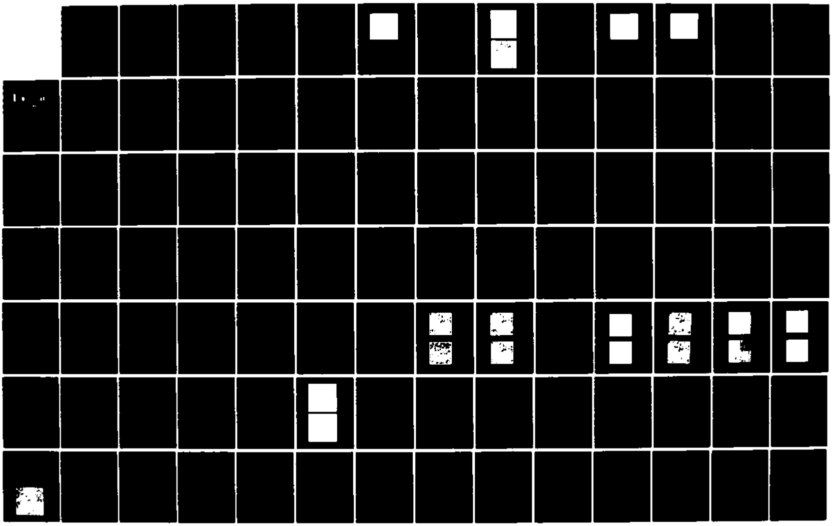
TABLE 5.2-1
COMPARISON OF CLASSIFICATION STRATEGIES

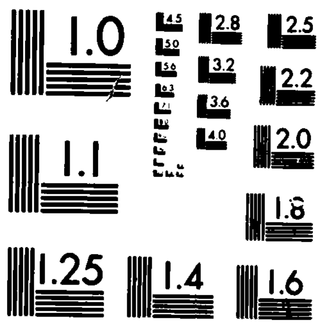
TECHNIQUE	GROUND TRUTH REQUIRED	COMPUTATIONAL COST	LEVEL-OF-AUTOMATION	PERFORMANCE
(1)	yes	medium	low	med
(2)	no	high	med	med
(3)	yes	low	low	high
(4)	no	medium	high	not available*

*Currently being determined by TASC.

AD-A150 189 FEATURE EXTRACTION ASSESSMENT STUDY(U) ANALYTIC SCIENCES CORP READING MA M J CARLOTTO ET AL. NOV 84 ETL-0377 DACA76-82-C-0004 3/4

UNCLASSIFIED F/G 8/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Of the above four strategies, the first is the least automated. It requires ground truth, and is moderately efficient. In the second, ground truth is not required. Training is replaced by signature matching, which can be automated, but should at least be verified by the user. Classification is performed on a pixel-by-pixel basis resulting in moderate classification performance (error rates) in (1), (2), and (4). In the third, performance can be controlled by varying the annexation threshold. Up to a 50% reduction in processing time has also been reported. The knowledge-based approach (4) promises the highest level of automation due to the generality of its knowledge. In addition, by structuring the classifier as a tree in (4), classification may be performed in a hierarchical fashion. That is, major classes such as water, soil-like materials, and vegetation are identified first. Each class is then decomposed into subclasses; e.g., vegetation into crops and trees, and soil-like materials into plowed fields, concrete, and silt. Sub-classification is performed until a specified level of detail has been achieved. Partial classification is also possible given incomplete data.

5.3 OBJECT IDENTIFICATION

Object identification is the process of inferring DMA features from the classification map. It involves grouping pixels with the same surface material type into regions, and computing attributes such as the size and shape of, and the relations between regions. It also involves grouping regions into objects (possible DMA features) using prior information about the types of materials likely to make up an object, as well as the structural and relational properties of the regions which constitute an object. Finally it involves inferring the identity of each object visible in the image based on the

attribute values of constituent regions. In the object identification scenario outlined in this chapter, machine processing is used to extract connected regions of the same surface material composition from the classification map, to compute structural attributes of the connected regions, to organize regions into groups based on relative attribute values, and to tentatively identify regions as DMA features. In the proposed scenario, machine processing is supervised, followed by human verification and editing.

The machine processing algorithms described below rely on two sources of information: the surface material classification map (the output of the surface material classification process), and contextual information (provided by the user or by other collateral data bases). It assumes that 2-d models of objects may be used for detection and initial identification. *Subsequent human verification and editing are required* to deal with situations in which unexpected objects/materials appear and/or the 2-d models are not adequate in cases of occlusion and severe perspective distortion.

Object identification involves the following major activities depicted in Fig. 5.3-1:

- Planning: Given a particular type of scene (agricultural, urban, residential), predict what objects are likely to be present. Possible objects constrain the types of materials likely to occur in the scene, as well as their structural and relational properties. This information is used to guide the interpretation process.

- Spatial Processing: spatially process (e.g., shrink/expand) binary images defining the location and extent of the various surface materials in the image to remove isolated pixels, extract compact

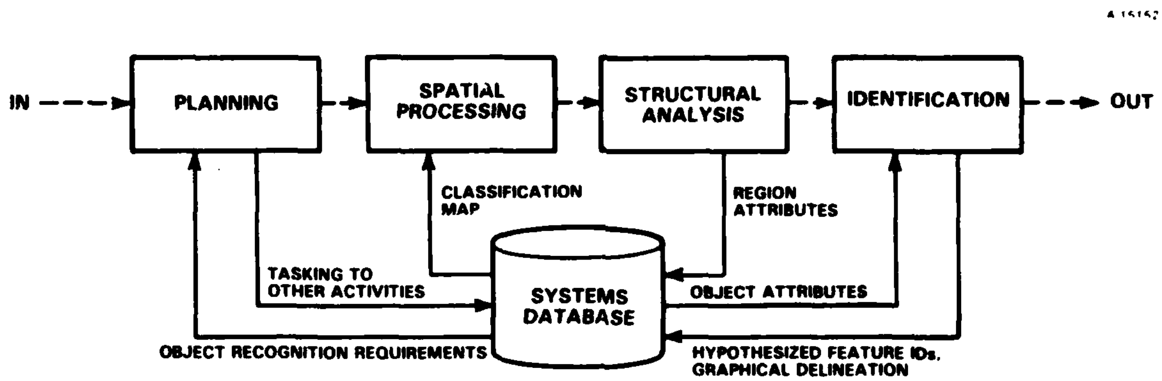


Figure 5.3-1   Object Recognition Activity-Flow

regions, merge adjacent regions, and so forth. Extract and label connected regions of similar material type.

- Structural Analysis: Compute attributes of connected regions; organize into groups according to relative attribute values (e.g., highly elongated regions composed of concrete)

- Identification: Using the results above, identify regions(s) as possible MC&G features using machine inference followed by human verification and editing.

The inputs to planning include collateral information such as the type of scene, location, time of day/year, and sensor parameters (spatial resolution, spectral characteristics), and a surface material classification map. Outputs from planning include an assessment of the scene (e.g., were unexpected materials found in the scene) for use in flagging situations in which human intervention may be required, direction to subsequent processes such as the types of attributes to be computed for what regions, and how this information can be used

to infer DMA features. Spatial processing operates on selected surface materials in the classification map. Structural analysis builds region attribute lists for use by man and machine in identification. Each activity is further detailed below.

5.3.1 Planning

Planning provides top-down guidance for both the object identification and surface material classification processes. The user supplies contextual information (kind of scene, location, time of day/year, sensor) from which information, such as the kinds of objects likely to be found in the scene, and in turn, the kinds of materials likely to make up these objects may be inferred. Such information could be passed down to other processes to help in the selection of classification and pre-processing techniques. Also, by comparing the kinds and relative proportions of surface materials found to those predicted, anomolous situations can be flagged and handled manually.

5.3.2 Spatial Processing

Spatial processing involves the extraction of selected surface material categories from the classification map, spatially processing the resulting occupancy arrays, and labeling connected regions for spatial referencing by subsequent activities. The input to this process is a surface material classification map (Fig. 5.3-2). The type of spatial processing that is required depends on image resolution, classification performance, and the eventual requirements of the identification process. For example, in lower-resolution imagery, thin objects such as roads may not form extended lines (a connected

Figure 5.3-2 Surface Material Classification Map. Color Code: Blue (water), Green (vegetation), Bright Green (crops), Brown (soil-like materials), Yellow (concrete/silt), and Red (plowed fields)

series of pixels labeled as concrete or another road-like material). As a result, line-growing may first have to be performed. If pixel classification is used, pixels along the border of large homogeneous regions (crop fields) may be misclassified. Such regions must be removed before further processing can take place. In searching for large agricultural areas in the image, nearby regions containing crops and plowed fields, for example, may be aggregated.

Within this activity the user thus has the following processing options:

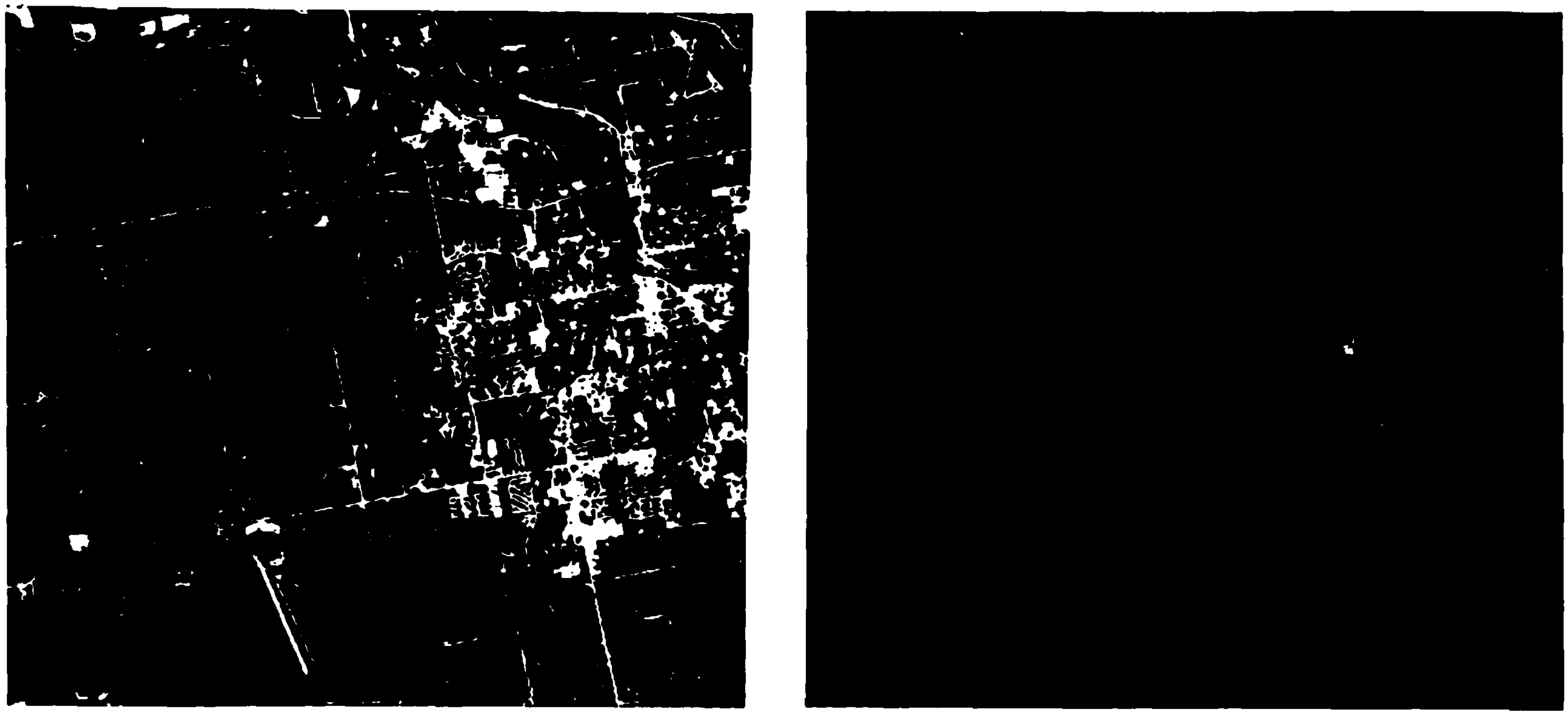
- Create a set of binary images, one for each surface material class
- Perform spatial processing. Options include region shrinking or expanding, as well as logical operations between binary images.
- Tag each of the remaining regions for referencing by the functions to follow.

As an example, regions having soil-like properties that are bright in the visible (concrete and silt) have been selected (Fig. 5.3-3a). Figure 5.3-3a is processed with a "shrink" operator to eliminate small and thin regions, followed by an expand operator to restore the regions which remained after shrinking to their original size. These large compact regions are colored yellow in Fig. 5.3-3b. By subtracting this image from Fig. 5.3-3a, the small and thin regions which were eliminated by the shrink/expand operation may be obtained. (These are colored green and red in Fig. 5.3-3b.)

5.3.3 Structural Analysis

Structural analysis involves computing the values of selected attributes such as

- Location (centroid of region)
- Area
- Perimeter
- Compactness (the area divided by the perimeter squared is often used)
- Elongatedness (e.g., the ratio of the maximum to minimum moments of inertia for each region)
- Orientation (the angle of the axis of least inertia)



(a) (b)

Figure 5.3-3 Example of Spatial Processing: (a) Pixels having soil-like characteristics that are bright in the visible; (b) Compact Regions (yellow), Elongated Regions (green), and Others (red)

of each connected region in the spatially processed surface material classification map. Region attributes are generally stored in some form of association or property list.

To aid both in manual interpretation and machine recognition, regions may be organized into groups based on the relative value of selected region attributes. For example, in Fig. 5.3-3b, non-compact regions composed of concrete/silt have been sub-divided into two groups: those that are elongated (green), and those that are not (red). An analyst (or machine algorithm) would recognize the former group as possible "road" segments, and would adjoin short intervening segments to form extended road network objects. Another example in Fig. 5.3-4 shows candidate agricultural regions, i.e., crops and plowed fields aggregated by expanding and shrinking crop and plowed-field regions, and sorted into three groups according to area: large (red), medium (green), and small (blue). This type of processing could be of use in determining where the major agricultural areas in an image are located.

5.3.4 Identification

Objects may appear to be composed of one region or more depending on the resolution of the imagery, as well as on the objects themselves. In Landsat TM imagery, while many roads can be detected manually, only the wider road-like regions can be detected as described in the preceding sections. However, as individual road regions (road surface, median strip, shoulders) become apparent at higher resolutions, grouping operations must take place to organize these kinds of regions into candidate road objects. Thus, as the resolution of the data increases, the complexity of object identification increases as well.

Figure 5.3-4 Agricultural Areas (i.e., crops and plowed fields) Sorted According to Size: Red (large), Green (medium), and Blue (small)

Assuming resolutions of the order of Landsat TM and Spot, many types of features can be identified on the basis of simple region properties alone. For illustration purposes only, bodies of water in Fig. 5.3-2 have been identified as ponds, rivers, and lakes. In Fig. 5.3-5 ponds (red) are defined to be regions composed of water, that are less than 25 pixels in area; lakes (green) must be greater than 25 pixels in area, but must have values less than 10 in elongatedness (roughly the length divided by the width); rivers (blue) must also be greater than 25 pixels in area, but must have length to width ratios roughly less than 10. These values are arbitrary, and again are used purely for illustrative purposes. An approach then for identifying a region(s) as an instance of

Figure 5.3-5 Bodies of Water Identified on the Basis of Two-Dimensional Region Properties: Lakes (green), River (blue), and Ponds (red)

a DLMS feature, for example, would be to translate selection requirements (e.g., SMC, minimum length/width) into rules which examine selected region attributes.

5.4 SUMMARY

Table 5.4-1 summarizes data processing requirements in the three major functional areas/activities within the feature extraction system. A large amount of image/numerical computation is performed during pre-processing for image registration, enhancement, and restoration. In semi-automated scenarios,

TABLE 5.4-1
SUMMARY OF DATA PROCESSING REQUIREMENTS

ACTIVITY AREA	DATA TYPE			
	GRAPHICAL	TEXTUAL	IMAGE/NUMERICAL	SYMBOLIC
Pre-Processing	low	med	high	low
Surface Material Classification	med	low	med	med
Object Recognition	med	low	low	high

little symbolic, textual, or graphical data is generated during pre-processing. As one proceeds into surface material classification and object recognition, an increasing amount of symbolic and graphical data processing is required since the imagery is being transformed into symbolic form and displayed in graphic form. As a result, numerical processing requirements decreases. Text processing is higher in pre-processing since it is responsible for overall system resource planning.

5.5 FUNCTIONAL ARCHITECTURE FOR MS/MS FEATURE EXTRACTION

This section outlines a potential functional architecture for a semi-automated MS/MS feature extraction system embodying the concept of operation and functional capabilities described in the previous sections. Figures 5.5-1 and 5.5-2 provide a block diagram and artist's conception of the architecture, which was originally introduced in Ref. 161. The similarities between the planned Remote Work Processing Facility (RWPF) Upgrade and the MS/MS feature extraction system architecture are several (e.g., VAX-class machine with array

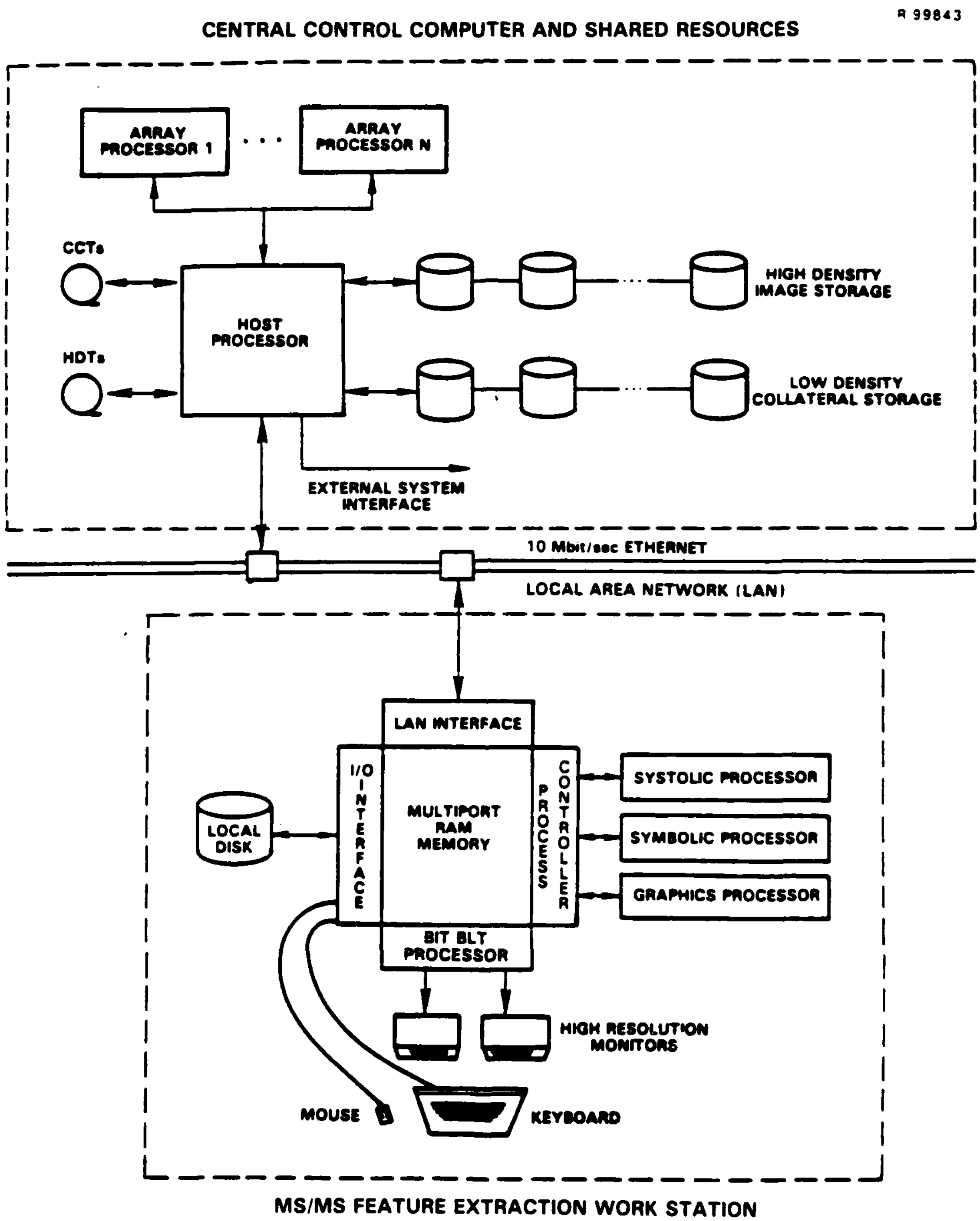


Figure 5.5-1 MS/MS Feature Extraction System Functional Architecture

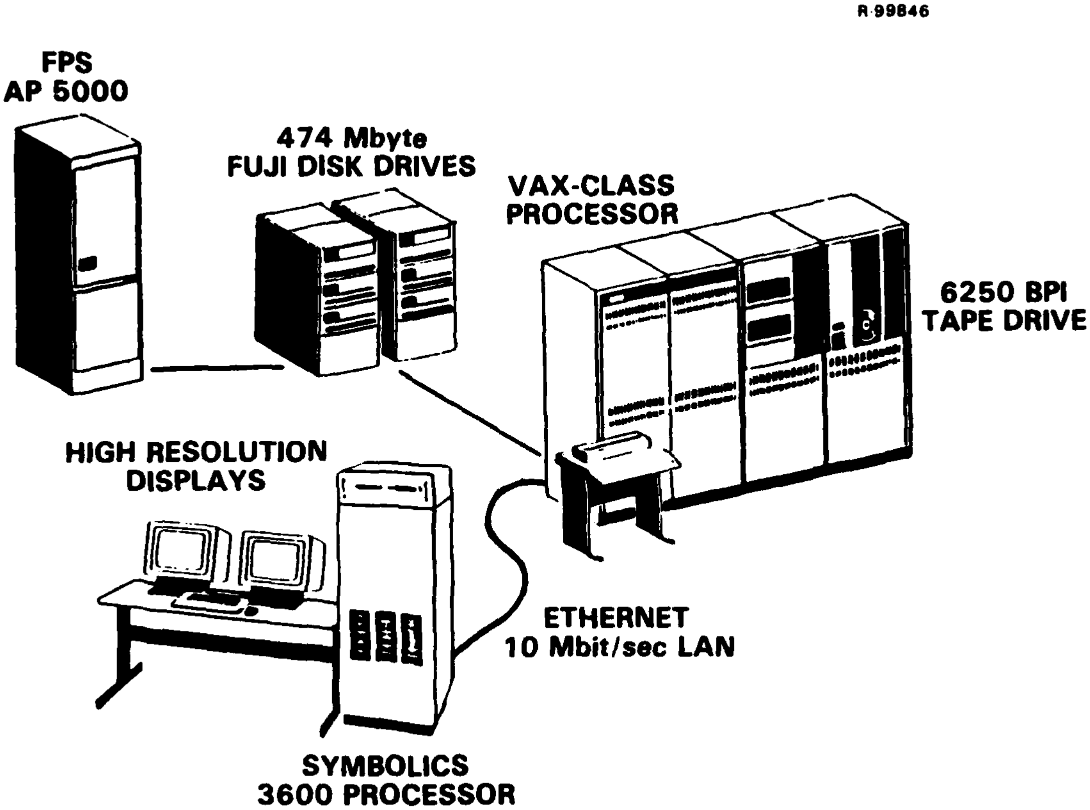


Figure 5.5-2 Artist's Concept

processor, large number of disks, Symbolics 3600-like workstations). The architecture portrayed here is intended for use in a production environment; as a result, certain elements of the system have been optimized accordingly (e.g., special-purpose processors attached directly to the workstation memory to improve I/O and function response times). In the sections below, the major elements of the system are discussed in more detail.

5.5.1 Central Control Computer and Shared Resources

The central control (CC) computer and its associated peripherals serve a number of functions within the MS/MS feature extraction system. First, the CC acts as the central file

server and database manager for all MS/MS feature extraction workstations supported by the system. In this capacity, the CC computer provides and manages all on-line and off-line storage for imagery, information derived from the imagery (e.g., pre-processed imagery, surface classification maps, region attribute files), collateral information (digitized maps and charts), and information needed to support the feature extraction process (e.g., rule bases for surface material classification and object recognition) through a combination of rapid access database management software (DBMS) resident on the host processor, high-density (e.g., optical disk-based) mass storage for imagery, and lower density (e.g., magnetic) storage for collateral data.

A second major function of the CC computer is to provide and manage all internal and external interfaces for the system. As a consequence, the CC computer must support interfaces between incoming digital imagery (e.g., seven track Landsat-TM data tapes) and collateral data (DFAD/DTED data tapes) in the form of computer compatable tapes (CCTs), high density tapes, or other media (e.g., optical disk) and the imagery/collateral data storage subsystem.

A third function the CC computer provides is to coordinate most of the numerically-intensive (image-level) processing in the system. Most of the functions performed during pre-processing, and many of the functions performed during surface material classification would be hosted on the CC computer. Among the processing alternatives capable of supporting the high numerical processing bandwidths required (typically greater than one megaflop), are attached array processors (FPS-5000, APTEC DPS-2400), parallel processors such as the TMI Connection Machine, and interconnected networks of high-speed VLSI-based processors (systolic arrays and special-purpose pipelined processors).

A final function the CC computer provides is a general system management function which ensures that all resources (man and machine) are optimally allocated based on the current workload, subsystem availabilities and internal/external I/O requirements. Thus the CC computer comprises both production management as well as computer resource management capabilities. System management functions also provide a processing environment that is functionally transparent to the user.

As can be seen, the demands upon the CC computer are diverse and potentially conflicting. Therefore, care must be exercised in specifying a host processor for the above system since it must provide adequate processing, communication, and storage capacity to support the above requirements. At this time, potential candidates for such a host system are Digital Equipment Corporation's VAX-11/780 (and larger machines), the Gould Concept 32/87, or Pyramid Technology's 90X machine.

5.5.2 MS/MS Feature Extraction Workstation

The MS/MS feature extraction workstation is the user's primary interface to the overall MS/MS feature extraction system, and supports all interactive processing within the system. Although intended to be both compact and relatively low cost, it encompasses a number of sophisticated functions. These functions correspond to four subsystems comprising the workstation: namely, the mass storage subsystem, local area network interface, processing subsystem, and man/machine interface.

The storage subsystem consists of two elements. The first is a local disk intended to function as a temporary imagery/collateral data working store. Data associated only with the currently active session or job resides on the disk and once the session is terminated, the data is written back

to the CC computer's storage subsystem or is expunged entirely (e.g., files containing intermediate results). In order to support the storage requirements associated with MS/MS imagery, a high density (at least one gigabyte) magnetic disk is recommended. The second element of the workstation storage subsystem (and a relatively unique element given current technology) is a high-speed multiport random access memory (RAM). This memory provides fast working storage for use by the other workstation subsystems as well as a shared communication path between all of the workstation subsystems. A multiport RAM consisting of between 10-20 megabytes of storage with access times less than 150 nanoseconds is envisioned.

The local area network (LAN) interface permits the workstation to communicate with the CC computer and other workstations via a moderate-to-high speed multi-access local area network. Such networks have become quite common and tend to center about either the Ethernet carrier sense multiple access, with collision detection (CSMA/CD) communication protocol or token passing ring network protocols. Since such a network within a MS/MS feature extraction system may have to support tens or even hundreds of workstations transferring imagery and collateral data, (at least) 10-50 megabit/sec data rates will be required. Therefore, the workstation's LAN interface must be designed to accommodate such data rates.

A third workstation subsystem is the processing subsystem. As shown in Fig. 5.5-1, this system consists of one or more special-purpose processors optimized to handle particular types of MS/MS feature extraction functions. In particular, a systolic processor would be desirable to provide real-time support of numerically-intensive image processing functions. A symbolic processor would be required to support higher-level

inferencing functions during object recognition and surface material classification. Finally a special-purpose graphics processor capable of creating and manipulating graphic overlays (e.g., maps and charts, surface material classification maps, and object icons) would also be desirable.

5.6 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

5.6.1 Summary

This chapter has described a concept of operation for an interactive, semi-automated MS/MS feature extraction system based on FY85 technology. After reviewing feature definitions, stating source imagery and collateral data assumptions, and specifying a baseline digital environment, the MS/MS feature extraction process was organized into three functional areas:

- Pre-Processing
- Surface Material Classification
- Object Recognition.

Pre-processing includes all activities relating to the preparation of MS/MS imagery for subsequent machine processing as well as those involving manual (i.e., computer-assisted) exploitation. Surface material classification is concerned with the determination of the physical composition of object surfaces visible in the imagery. In object recognition the surface material classification map and 2-d attributes derived from it are used to infer the identity of features visible in the image.

Initially, generic control/data-flows within and between the above functional areas were identified for the purpose of elucidating an end-to-end process-flow within the system. The activities performed within each functional area were then described. Examples were presented illustrating the following activities within the MS/MS feature extraction process:

- Image Registration: Registration of Thematic Mapper and Seasat-SAR imagery using an automatic control point selection procedure.
- Color Enhancement and Display: Use of a local histogram equalization technique to enhance planimetric features in Landsat MSS imagery.
- Landsat MSS Image Classification: Example illustrating a supervised classification of Landsat MSS imagery into major land-cover classes. An example of how class and confidence (i.e., how well the classifier is performing) may be displayed together in color was also presented.
- Object Recognition: Examples of various activities performed during object recognition including spatial processing, structural analysis, and identification were included

Also within various activities, alternate techniques were assessed. In particular, four classification strategies were compared with respect to computational cost, performance, and level-of-automation possible.

5.6.2 Conclusions

The conclusions of this chapter are two-fold. First with respect to technical feasibility, it is shown that use of MS/MS imagery can significantly increase the level-of-automation

possible in an FY85 feature extraction system. In particular, surface material classification, established as a key processing step in the MS/MS feature extraction process, can be automated to a considerable extent given current pattern recognition technologies. Additional improvements should be possible using knowledge-based techniques. Second, with regard to current sensor systems (e.g., Landsat TM), low spatial resolutions (30 m typical) will limit their use in the compilation/updating of large scale maps. Their spectral resolution, however, appears to be adequate for extracting surface material classes of interest to the cartographer (e.g., vegetation and soils, concrete and other road materials, and water).

5.6.3 Directions for Further Research

Finally, with respect to potentially beneficial areas for experimentation and research, it is recommended that efforts be devoted to:

- Determining a minimal set of 2-d attributes for identifying representative DMA features in imagery acquired by a particular sensor (e.g., Landsat TM). This can later be extended to other sensors and combinations of sensors.
- Developing an improved testbed capability for conducting experiments to quantify the performance of techniques identified by this study (the RWPF-upgrade would be a likely system to host such a testbed).

Only after these efforts are completed, can the real utility of MS/MS imagery and the cost effectiveness of a semi-automated MS/MS feature extraction system be determined.

6. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

This report presents an assessment of image processing, pattern recognition, and artificial intelligence techniques of potential use in increasing possible automation in a FY85 feature extraction system. It also provides descriptions of how these techniques may be used operationally at DMA for extracting features from black and white and multi-spectral/ multi-source imagery. This chapter summarizes this work, states the major conclusions, and presents recommendations for further research.

6.1 BLACK-AND-WHITE FEATURE EXTRACTION TECHNIQUES

6.1.1 Summary

Chapter 2 has reviewed and assessed the current state-of-the-art in black-and-white feature extraction technology as it applies to the global DMA feature extraction process. Following a brief overview of candidate approaches for performing technique assessment, a particular approach - the IU paradigm - was described and its application to technique assessment discussed. The IU paradigm evaluation technique was then applied to several major classes of feature extraction techniques including edge extraction, segmentation, texture, statistical and syntactic pattern recognition and symbolic matching.

6.1.2 Conclusions

The approach selected for assessing candidate feature extraction techniques is to analyze the general machine visual-perception problem in terms of the IU paradigm. This paradigm is a model for showing how the data within an image must be interpreted (in terms of its relationship to the physical properties of the objects it portrays) in order to permit the accurate inference of descriptive object properties. It also shows that the ability to solve complex recognition problems like feature identification hinges on the ability to infer such object properties reliably.

The IU paradigm focused on the fact that local surface information, the information that fundamentally characterizes object appearance, is lost in the image formation process. Thus, recovery of physical properties must therefore proceed on the basis of global image properties. However, it was shown that to infer conclusively an object property from a global image property requires severe image acquisition, object orientation and property property constraints. These constraints had particularly strong impact upon region-based segmentation techniques. Since many of the latter techniques find regions which are characterized by some measure of signal uniformity, yet there is no clear relationship between such measures and actual object properties. However, while image regions appear to have no clear relationship to feature properties, image edges may be generated by physical phenomena of interest, including object boundaries. As a consequence, edge extraction techniques appear promising for the delineation problem. However, since image edges may not be well-defined due to the high level of detail of aerial imagery and its discrete nature, determination of the "best" edge extraction technique(s) to apply can only be accomplished within the scope of an operational scenario.

Also addressed were texture, pattern recognition and symbolic methods. Textures did not appear to be sufficiently robust due to their intrinsic statistical characterization of regions which seldom are satisfied in reality. Pattern recognition techniques suffered from the same weakness. Symbolic techniques do appear promising, but are not currently useable because they rely heavily upon lower-level edge extraction and segmentation techniques.

Finally, it was observed that an important requirement for machine perception is perceptual organization, the problem of choosing which local portions of an image belong to the same object or feature. It was also observed that this problem is similar to the feature delineation problem, for which it is not necessary to identify the semantic class of a feature in order to find its boundary.

6.1.3 Directions for Further Research

The recurring theme throughout Chapter 2 was that low-level techniques are the weak link in the machine perception hierarchy. There has been recent promising work in several areas, however, which indicates progress towards overcoming those weaknesses.

The first area involves relating image information to the nature of object properties as they are instantiated via the image formation process. Witkin (Ref. 35) showed that the behavior of the autocorrelation function of a window as it slides across on edge boundary may indicate whether the boundary is an occlusion or a shadow. The autocorrelation function across a shadow boundary will generally be smooth, because only the mean intensity changes and is averaged out by the correlator. The partially-shadowed region is likely to be otherwise homogeneous.

A second example is an attempt to define a texture representation which is physically and mathematically justified. Pentland (Ref. 36) showed how a mathematical function for randomness is related to the non-deterministic behavior of the undersampling of objects in aerial images, which creates detail phenomena.

The second area focuses specifically on the problem of perceptual organization. Lowe and Binford (Ref. 37) suggest that recognition is likely to be intimately involved with low-level perceptual organization, and propose general principles that govern the grouping process. Fischler (Ref. 38) has developed an image line-finding algorithm which is claimed to be capable of finding image lines (not image contour locations which correspond to physical edge phenomena) as well as humans can. It should be emphasized that this comparison is best made when the lines appear to be without familiar context, so that human viewers do not make knowledge-based interpretations.

The research directions described above are consistent with the needs indicated in this chapter, that low-level techniques must address problems of perceptual organization, and that meaningful low-level techniques should relate image properties to the instantiated physical properties of objects. Further emphasis and support of these research directions would appear to be particularly beneficial.

6.2 CONCEPT OF OPERATION FOR A SEMI-AUTOMATED FEATURE EXTRACTION SYSTEM

6.2.1 Summary

Chapter 3 outlined a concept of operation for an interactive, semi-automated feature extraction system based on

current technology. First, a generic concept of operation for feature extraction was described. Activities within the concept of operation were then identified which appear to be candidates for automation and the application of machine perception. The particular activities selected for automation were feature detection, identification and delineation, and based on the results of Task 3, alternative methods for applying machine perception were proposed. Subsequently, the feasibility issues and cost/benefit trade-offs surrounding the alternative methods proposed were discussed. Based on the results of the latter analysis, a refined concept of operation for a semi-automated feature extraction system was provided.

6.2.2 Conclusions

With respect to technical feasibility, it was argued that neither source data characteristics (e.g., resolution, scale variation, noise, solar elevation/azimuth) nor digital workstation technology (e.g., storage, processing, communication and displays) posed any feasibility constraints on the implementation of a semi-automated feature extraction system. However, the limitations of current feature extraction techniques and machine perception technology were felt to constitute a relatively high technical risk. In order to reduce this risk, the concept of a minimal attribute set of a feature was defined which would provide necessary but not sufficient information for the identification of a feature from strictly image-derived attributes. The objective of the MAS was to provide a capability which with high reliability would determine when and where no features of interest were present in an image, and provide cues to the possible location of features where MAS conditions were satisfied. Although the MAS concept substantially improves the potential feasibility of a semi-automated system, whether or not a MAS can be defined for each

feature or class of features to be extracted is currently an open problem.

With respect to the cost/benefit tradeoffs associated with the concept of operation proposed, it was argued that in each of the major trade-off categories - machine time vs manual time to perform a given task, man/machine interaction efficiency, equipment vs labor costs, throughput and accuracy - a semi-automated capability was superior to a purely manual capability. However, the tradeoffs can be quantified only in the context of a particular system implementation concept and in conjunction with experiments that would be designed to determine the relative performance of machine perception versus human interpretation in performing selected feature extraction tasks.

6.2.3 Directions for Further Research

Finally, with respect to potentially beneficial areas for experimentation and research, it is recommended that efforts to be devoted to:

- determining the feasibility and defining the characteristics of the minimal attribute set for a number of features of interest
- developing an improved testbed capability for hosting, in a more realistic production environment, experiments to quantify the relative performance of the semi-automated capabilities proposed versus manual capabilities
- defining how the man/machine interface should be developed for this system so that synergistic, rather than conflicting, interaction between man and machine can be realized.

Only after these efforts are completed can the real feasibility and cost/benefit of a semi-automated feature extraction system be determined.

6.3 MULTI-SPECTRAL/MULTI-SOURCE FEATURE EXTRACTION

6.3.1 Summary

Chapter 4 reviewed representative multi-spectral and synthetic aperture radar (SAR) sensors, and ssessed the use of image processing, segmentation, classification, and object recognition techniques in exploiting the data provided by thse sensors for feature extraction.

6.3.2 Conclusions

Of the four sensor systems reviewed: LANDSAT, SPOT, SEASAT SAR, and the Shuttle Imaging Radar (SIR), the LANDSAT MSS is the only multi-spectral sensor system to have achieved operational status; the other sensors were largely developmental in nature. However, a spatial resolution of 56 × 79 m in the visible and reflective IR limits the use of the MSS in feature extraction applications. The TM has both superior resolution (30 m visible and reflective IR), and spectral resolution (7 bands, including thermal IR) over the MSS. The added bands gives the TM improved ability to discriminate geologic resources, types of vegetation, and land use. However, its resolution still limits its use to compiling and updating large scale maps only. The planned French SPOT satellite will have better spatial resolution (20 m multi-spectral and 10 m panchromatic), but will have fewer spectral bands (only 3).

The SAR sensors (SEASAT SAR, SIR-series) reviewed are all considered to be experimental in nature. All are L-band radars (23.5 cm) with a spatial resolution of about 25 m (four looks averaged). Both the SEASAT SAR and SIR-A were uncalibrated devices. In exploiting this type of SAR imagery one must be aware of possible limitations in dynamic range of the data (typically 4 bits), and must be prepared to deal with a considerable amount of "speckle". Although comparable in spatial resolution to the TM, cartographic accuracies for these sensors are lower, at least ± 50 m (and possibly less, depending on how careful the user is in selecting control point pairs).

Among the image classification techniques discussed in this section, pixel classifiers are the simplest to design and implement, but are not very efficient. Since region classifiers process groups of pixels at a time, depending on how expensive it is to group pixels into regions, region classification can be quite efficient (up to a 50% decrease in classification time has been reported in Ref. 130). Since region classifiers make use of information from neighboring pixels, classification accuracy can also be improved. Multi-temporal technqiues increase the ability to discriminate between, and classify certain types of vegetation. Signature extension allows the spectral signatures of known material types in one image to be mapped to another. Since all of the above technqiues require some degree of supervision (training or signature mapping), the degree to which the surface material classification process can be automated is limited. Knowledge-based techniques have the potential to further automate the process, however, additional work is required.

Of all MS/MS technology areas, image processing appeared to be the most mature. Past work in remote sensing has provided many techniques of potential use in feature extraction.

In some cases, techniues that were originally developed for optical imagery are directly applicable to multi-spectral and multi-source imagery. For example, geometric transform techniques developed for optical imagery are useful for registering SAR and multi-spectral data sets as well. On the other hand, while black-and-white (single image) enhancement techniques can be applied on a band-by-band basis, new techniques which exploit correlations between bands (for thermal band sharpening) and between sensors (for using coregistered optical imagery to smooth SAR) appear promising.

Finally, new image transformations based on the tasseled cap/canonical variates approach which provide physically-significant information (e.g., vegetative cover, soil moisture) can be expected to be of considerable utility to the image anlaysis in manual interpretation as well as in surface material classification.

Several computer vision systems were assessed as candidate automatic MS/MS feature extraction systems. Systems developed at Kyoto University (Ref. 110) and the University of Massachusetts (Ref. 26), which use 2-d models to represent objects of interest in a scene, appeared applicable. Although additional developments in this area will be necessary before an automatic system can be developed, the 2-d approach did appear promising for feature extraction. Such an approach has been shown to be applicable in aerial imaging application where the illumination is far from the scene, the view angle is relatively fixed over the field of view, and occlusion is not a significant factor.

6.3.3 Directions for Further Research

While image pre-processing is considered to be a fairly mature technology area, further work in several areas is recommended. First, our assessment of image restoration and enhancement techniques revealed that while many single-band (monoscopic) techniques exist, few make explicit use of more than one band or sensor. Initial results presented in Appendix F demonstrated the utility of multi-band/sensor techniques for spatial enhancement. It is recommended that additional work be performed to quantify the performance of multi-band thermal band sharpening and SAR smoothing techniques, and to investigate other applications of the technique. (One such use for detecting and restoring data drop-outs was suggested in the report.) The use of tasseled cap transforms and canonical variates to extract information such as vegetative cover, wetness, and concreteness from an image, should also be pursued. In particular, transforms for other physically-significant properties should be derived.

Alternate image classification strategies (e.g., knowledge-based techniques as described in Appendix G) need to be more fully developed and tested. An operational assessment of different image classification strategies (with ground truth) should be conducted to determine the merits of heuristic versus statistical techniques (i.e., to what extent can heuristic rules increase the level of automation possible in the classification process), to determine to what extent region-based classification is superior to pixel-based classification (e.g., in terms of error rate, and processing time), and to determine to what extent prior information (e.g., context) improves classifier accuracy. The assessment should be performed using a variety of scenes (agricultural, residential, and urban), acquired at different times (time of day and season), and under a variety of scene/sensor conditions (haze, sensor noise levels).

Finally, it is suggested that a testbed be assembled for assessing MS/MS feature extraction techniques. (The RWPF-upgrade would be a candidate target system.) The objectives of the testbed would be three-fold: to allow experimentation with diverse imagery sources to determine what kinds of information can be readily extracted from what types of imagery under what conditions, to allow prototypical feature extraction systems (i.e., special-purpose vision systems) to be developed and tested, and to provide an environment for DMA to transition new these new feature extraction technologies into production systems.

6.4 CONCEPT OF OPERATION FOR A MS/MS FEATURE EXTRACTION SYSTEM

6.4.1 Summary

Chapter 5 described a concept of operation for an interactive, semi-automated MS/MS feature extraction system based on FY85 technology. After reviewing feature definitions, stating source imagery and collateral data assumptions, and specifying a baseline digital environment, the MS/MS feature extraction process was organized into three functional areas:

- Pre-Processing
- Surface Material Classification
- Object Recognition.

Pre-processing includes all activities relating to the preparation of MS/MS imagery for subsequent machine processing as well as those involving manual (i.e., computer-assisted) exploitation. Surface material classification is concerned with

the determination of the physical composition of object surfaces visible in the imagery. In object recognition the surface material classification map and 2-d attributes derived from it are used to infer the identity of features visible in the image.

Initially, generic control/data-flows within and between the above functional areas were identified for the purpose of elucidating an end-to-end process-flow within the system. The activities performed within each functional area were then described. Examples were presented illustrating the following activities within the MS/MS feature extraction process:

- Image Registration: Registration of Thematic Mapper and Seasat-SAR imagery using an automatic control point selection procedure.
- Color Enhancement and Display: Use of a local histogram equalization technique to enhance planimetric features in Landsat MSS imagery.
- Landsat MSS Image Classification: Example illustrating a supervised classification of Landsat MSS imagery into major land-cover classes. An example of how class and confidence (i.e., how well the classifier is performing) may be displayed together in color was also presented.
- Object Recognition: Examples of various activities performed during object recognition including spatial processing, structural analysis, and identification were included

Also within various activities, alternate techniques were assessed. In particular, four classification strategies were compared with respect to computational cost, performance, and level-of-automation possible.

6.4.2 Conclusions

The conclusions of Chapter 5 are two-fold. First with respect to technical feasibility, it is shown that use of MS/MS imagery can significantly increase the level-of-automation possible in an FY85 feature extraction system. In particular, surface material classification, established as a key processing step in the MS/MS feature extraction process, can be automated to a considerable extent given current pattern recognition technologies. Additional improvements should be possible using knowledge-based techniques. Second, with regard to current sensor systems (e.g., Landsat TM), low spatial resolutions (30 m typical) will limit their use in the compilation/updating of large scale maps. Their spectral resolution, however, appears to be adequate for extracting surface material classes of interest to the cartographer (e.g., vegetation and soils, concrete and other road materials, and water).

6.4.3 Directions for Further Research

Finally, with respect to potentially beneficial areas for experimentation and research, it is recommended that efforts be devoted to:

- Determining a minimal set of 2-d attributes for identifying representative DMA features in imagery acquired by a particular sensor (e.g., Landsat TM). This can later be extended to other sensors and combinations of sensors.
- Developing an improved testbed capability for conducting experiments to quantify the performance of techniques identified by this study (the RWPF-upgrade would be a likely system to host such a testbed).

Only after these efforts are completed, can the real utility of MS/MS imagery and the cost effectiveness of a semi-automated MS/MS feature extraction system be determined.

APPENDIX A
IMAGE ENHANCEMENT

This appendix describes techniques that are used to enhance operational imagery for the purpose of facilitating interactive feature extraction. The three main classes of techniques examined include: contrast enhancement techniques, sharpening and edge enhancement techniques, and noise suppression techniques. Tables A-1 through A-3 summarize representative techniques within each class.

A.1 CONTRAST ENHANCEMENT

Standard contrast enhancement techniques (which include the first three categories in Table A-1) are covered in (Refs. 9, 41-51). Most of these techniques have already been implemented in fast display hardware. Adaptive techniques, however, are more recent developments and require much more computation. Two examples of the latter techniques are given by Peli (Ref. 52), who describes how adaptive contrast stretching can be used for removing haze in aerial photography, and Tom (Ref. 53), who developed an optimal transformation based on maximum entropy for bringing out as much information from imagery as possible.

Geometric remapping is a major consideration in spatial enhancement. A general description of geometric remapping for digital imagery is provided in Refs. 9, 41-43, 45. Nearest neighbor, bilinear, cubic convolution, and cubic spline, are standard techniques for achieving remapping. Of the first

TABLE A-1

SUMMARY OF CONTRAST ENHANCEMENT TECHNIQUES

GENERAL CLASS	REPRESENTATIVE ALG.	SUMMARY	COMMENTS
Fixed transformations (single parameter)	Contrast stretch	Computes intensity level remappings based on simple models. Linear stretch is the simplest of this group. Other forms take into account human visual perception and are modified log mappings.	Fast and efficient way to enhance predominantly contrast limited images.
Interactive transformation (multiple parameters)	Rubber-band	Specifies the output intensity mapping via an interactive trackball program. This provides the most general form for the intensity remapping function.	A convenient way to enhance contrast limited imagery. More interaction is required
Global data dependent transformations	Clip & stretch, histogram specification	The histogram of the input imagery is computed to determine the output remapping function. For clip & stretch, the cutoff intensity levels are specified giving a percentage of the histogram to cut off. For histogram specification, the *remapping* function is chosen so as to map the input histogram to the desired histogram form.	Less interaction is required since these methods use the input data to compute the best remapping function for displaying the data.
Locally data dependent	Adaptive contrast stretch, adaptive histogram equalization	These techniques are based on sliding windows which compute local information (mean, variance, or histograms) and use this information to optimally enhance the pixel which is centered in the window.	Best blind enhancement techniques. The amount of sophisticated computation required is proportional to number of image pixels.

TABLE A-2
SUMMARY OF SHARPENING AND EDGE ENHANCEMENT TECHNIQUES

GENERAL CLASS	REPRESENTATIVE ALG.	SUMMARY	COMMENTS
1st-difference/ direction sensitive	Sobel, Compass Gradient	Computes the approximation to a gradient, i.e., first difference using 3×3 finite masks. The techniques in this group use masks of several orientations. Some of the techniques are nonlinear, i.e., involve max, absolute values, or square magnitude to generate an edge output. Orientation and magnitude can be generated.	Fast and efficient way to enhance dominant image edges on noise-free imagery.
2nd-difference/ direction sensitive	Laplacian	Computes the approximation to a Laplacian, i.e., 2nd difference using 3×3 finite masks. Only magnitude information is generated.	Results similar to above. Susceptible to high frequency noise.
Edge-fitting	Hueckel	Computes the edge image by solving a minimum "distance" problem using ideal edge masks of varying orientations.	Less susceptible to noise. Edge orientations quantized.
Large mask	Argyle	Split Gaussian weighting of an approximate first derivative operator. The Gaussian weighting affords some local smoothing on each side of the edge.	Gaussian weighting provides some noise immunity and emphasis on the border.
Frequency Domain	V shape, Prolate, Wiener	2-D FFT based filter using circularly symmetric "conic" shapes for optimal edge enhancement. Yields closer approximations to ideal deriv. function. The equivalent FIR mask is proportional to the size of the FFT. V filter is adequate for noise-free imagery.	Can be done locally adaptively using short-space overlapped blocks. Very general filter format for trading sharpening for noise smoothing.
Multiple-resolutions	Marr-Hildreth, Rosenfeld-Thurston	Computes edge images at different resolutions of the source imagery by using different mask sizes. The Marr-Hildreth uses circularly symmetric 2nd derivative Laplacian operators.	Different resolutions are used to confirm the significance of edges.

TABLE A-3

SUMMARY OF NOISE SUPPRESSION TECHNIQUES

GENERAL CLASS	REPRESENTATIVE ALG.	SUMMARY	COMMENTS
Linear	Low-pass mean	Smooths the data by convolving the data with finite mask sizes. Can be implemented in the frequency domain as magnitude filter (Lowpass, Weiner).	Reduces noise at the sacrifice of smoothing the data. Combined with a drop-out detector, techniques are more useful.
Non-linear	Median	Does a nonlinear computation on a sliding window (finite mask size) such as a median to compute the filtered output.	Reduces isolated pixel values which are noise. Preserves sharpness of features better than above.
Spatially adapting	Edge-preserving smoothing	Uses sliding multiple masks of varying orientations and computes local homogeneity measures in order to choose the best mask orientation for filtering.	Simplifies imagery as well as reducing noise. Will tend to aggregate areas of similar intensity.

three techniques, bilinear interpolation has been shown to be superior for operational imagery, since it produces less staircasing then nearest neighbor interpolation and less blurring than cubic convolution interpolation (Ref. 2). It is less clear whether bilinear is better than cubic spline. Crochiere (Ref. 75) developed a resampling technique which assumes that the image data is bandlimited and can be implemented either in the spatial or frequency domain. This technique uses all the data to compute each interpolated output point.

A.2 SHARPENING AND EDGE ENHANCEMENT

Sharpening and edge enhancement are the other main considerations in spatial enhancement. Techniques for image sharpening have been addressed in Refs. 9, 14, 41, 42, 51, 54-60. The primary method for sharpening has been high frequency emphasis (HFE) filtering (Refs. 14, 56-58). Schreiber sharpened images by adding over and under shoot to feature boundaries using a photogrammetric technique called unsharp masking (Ref. 59). This method can be shown to be equivalent to the HFE filter. Stockham applied an HFE filter in the log domain in an attempt to separate reflectance effects from illumination effects (Ref. 60). Schreiber unified these concepts in Ref. 50.

Edge enhancement is typically accomplished by using one of two techniques: spatial differentiation or edge fitting by approximate models. Reviews of the performance of elementary edge operators are given in Refs. 21, 61-65. Edge operators that are based on derivative functions comprise virtually all of the techniques in Table A-2 with exception of edge-fitting. Examples of such operators are provided in Refs. 12, 14, 15, 19, 22, 66, 70-74. The Hueckel edge detector is a good example

of an edge-fitting method (Refs. 68, 69). Methods that utilize these operators at varying resolutions to reinforce the enhancement of significant edges were described by Rosenfield (Ref. 67) and Marr (Ref. 15).

A.3 NOISE FILTERING

Noise filtering techniques that appear to be applicable to feature extraction fall into three categories: linear, non-linear, and spatially adaptive techniques. Linear and non-linear (e.g., median) filtering techniques are covered in Refs. 9, 41, 42. The specific properties of median filters have been characterized by Tyan (Ref. 78). Spatially adaptive techniques are described in Refs. 76, 77, 79, 80.

APPENDIX B
EDGE THINNING AND LINKING

This appendix describes techniques that are used to process raw edges that have been detected within an image, and produce usable line or feature boundary descriptions suitable for extraction or delineation. These techniques are called line thinning (or skeletonizing) and line linking techniques. These procedures are not robust and should be considered within an interactive environment only.

B.1 EDGE THINNING

A general description of the heuristic nature of thinning can be found in Ref. 42. Most thinning algorithms operate on binary (black-and-white) images, but Dyer (Ref. 81) has done some work on grey level images. One common method used to skeletonize regions is the medial axis transformation (MAT) described by Blum (Ref. 82). In the ideal, non-discrete image case, the MAT is unique and invertible. In the discrete pixel case, however, there are problems in the implementation of the MAT. Montinari (Ref. 83) performs a digital polygon approximation before applying the MAT. Yokoi et al. (Ref. 84) fit discrete disks of varying sizes inside the regions of interest to compute the MAT using the center points of the disk positions. Pavlidis (Ref. 85) has come up with an implementation of the MAT that can be efficiently run on parallel processors. An efficient implementation of skeletonizing based on equidistance criteria from boundaries was performed by Arcelli (Ref. 88). Other approaches to thinning apply the concept of

local connectivity (e.g., Pfaltz, Ref. 86). A review of thinning based on connectivity can be found in Ref. 87.

B.2 EDGE LINKING

There are basically three major classes of line segment linking techniques: Hough transform methods (Refs. 20, 89, 90), sequential tracking methods (Refs. 23, 91, 92), and parallel propagation or connectivity methods (Refs. 93, 94). Hough transform methods map line data to a feature space so that co-linear points accumulate in isolated bins. Sequential tracking methods start on line contours and perform a directed search from the starting position using heuristic rules to follow the contour. Parallel methods perform local aggregation and iteration on the line segments. The general grouping of edge points into higher order entities is reviewed in Ref. 42. More experimental work on operational imagery needs to be performed before any of these techniques can be recommended for use in a production interactive feature extraction workstation.

APPENDIX C
SEGMENTATION

This appendix describes several major classes of segmentation techniques, including: region growing, region splitting, split-merge, thresholding and clustering. Representative techniques from the latter classes of segmentation algorithms are summarized in Table C-1.

C.1 REGION GROWING

Region growing merges fundamental, "atomic" regions into larger regions based on textural and/or radiometric similarity. In an early algorithm (Ref. 96), regions are merged if their intensities are similar. Brice and Fennema (Ref. 27) use a similar approach, but perform additional grouping across weak boundaries if the new region has a smaller boundary. This also removes small islands in the segmentation. The performance of these algorithms is dependent on the value of a global threshold. Nagao and Matsuyama (Ref. 79) compute a global threshold based on a histogram analysis of the gradient image.

There is currently disagreement on whether semantic information should be incorporated into the segmentation process. Semantically guided algorithms (Ref. 97) attempt to arrive at a globally consistent interpretation by assigning labels and probabilities to regions. The labels and probabilities are updated until the probability of a particular global interpretation is maximized, given the original measurements and context.

TABLE C-1
SUMMARY OF SEGMENTATION ALGORITHMS

GENERAL CLASS	SAMPLE TECHNIQUE (AUTHOR)	SUMMARY	COMMENTS	APPLICATION AREA
Region Growing	Muerle & Allen	• Join region if statistically similar	• Results dependent on threshold value	Scene Analysis
	Brice & Fennema	• Partition image into "atomic regions" • Join atomic regions if boundary is weak and new boundary is shorter • Smooth boundaries	• Results dependent on threshold value and merging constraints	Scene Analysis
	Nagao & Matsuyama	• Smooth image using edge-preserving smoothing algorithm • Compute global threshold from gradient image • Merge pixels if difference is less than threshold • Combine small segments with larger neighboring segments having similar spectral properties	• Threshold computed from image	Multi-spectral image interpretation
Region-Splitting	Ohlander	• Recursively split regions using thresholds computed from regional histograms • Continue until histograms in all segments for all features and uni-modal	• Uses multi-spectral and textural features • May be used to obtain a partial segmentation	Scene Analysis
Split-Merge	Pavlidis & Horowitz	• Split regions that are inhomogeneous • Merge regions that are similar	• Uses pyramidal data structure	Radiography and FLIR Image Segmentation
Thresholding	Chow & Kaneko	• Compute maximum likelihood threshold over 50% overlapping blocks in image • Interpolate threshold to all points in the image and threshold	• For 2-class object/background separation • Insensitive to variations in illumination	Radiography
Clustering	Coleman	• Clusters data using version of K-means algorithm • Number of classes (K) is increased until overall cluster quality is maximized	• Attempts to find the intrinsic number of clusters in the data	Scene Analysis

C.2 REGION SPLITTING

Recursive region splitting (Ref. 30) selects a region (initially, the entire image), computes histograms over its features (e.g, color, texture) and establishes a threshold. The threshold is used to partition the region into two or more subregions, each of which is split, recursively. This technique terminates when all regions are homogenous (uni-modal) across all features. The advantage of splitting over region growing is that it may be used to obtain a partial (coarse) segmentation of an image with very little effort.

C.3 SPLIT-MERGE

Horowitz and Pavlidis (Ref. 98) developed a split-merge approach for segmentation based on regional approximation. They construct a grey-level pyramid of reduced resolution images where the value of a father node in the associated tree structure is equal to the average of four sons. Regions with similar approximation are merged while regions with large approximation errors are split.

C.4 THRESHOLDING

Chow and Kaneko (Ref. 99) computed local thresholds in order to segment x-ray images into two classes (object and background). Histograms are first computed in 50% overlapping windows. Maximum likelihood thresholds are estimated assuming the object and background densities are Gaussian. The thresholds are then interpolated for all points in the image.

C.5 CLUSTERING

Image segmentation is essentially a unsupervised classification or clustering problem where, typically, the number of clusters and their properties are unknown. Coleman (Ref. 31) developed an iterative scheme which begins by clustering the data into two classes. A measure of cluster quality which incorporates intercluster separation and intra-cluster compactness in then performed. Clustering is repeated, increasing the number of classes each time, until the cluster quality satisfies an a priori value.

APPENDIX D
TEXTURE

This appendix describes several major classes of texture measurement techniques, including: co-occurrence matrices, linear filtering models, local histograms, power spectrum and structural techniques. Techniques that represent and measure texture are summarized in Table D-1.

D.1 CO-OCCURRENCE MATRIX

The co-occurrence matrix (Ref. 34) is a 2-d histogram of joint grey-levels taken at two points spaced a fixed distance apart in the image. The co-occurrence matrix is the basis from which various textural features can be computed (e.g., contrast, homogeneity, and correlation). Davis and Rosenfeld (Ref. 80) has generalized the above idea to include measurements derived from the original image (e.g., edges) and general predicates which describe relations between edges such as orthogonality. The co-occurrence matrix is probably the most popular and successful technique for texture representation.

D.2 LINEAR FILTERING MODELS

A 2-d auto-regressive moving-average (ARMA) model for texture generation and detection (Ref. 100) is an example of a linear (recursive) filtering model for texture. By assuming that texture can be synthesized by driving an ARMA model with white noise, the detection can be formulated as a hypothesis

TABLE D-1
SUMMARY OF TEXTURE REPRESENTATION

GENERAL CLASS	PARTICULAR TECHNIQUE (AUTHOR)	SUMMARY	COMMENTS
Co-occurrence Matrix	Haralick et al.	• Joint histogram of image grey-levels computed at two points a fixed distance apart • 2-nd order statistics of co-occurrence matrix used as textural features	• Most popular method • Applied to diverse problems ranging from photo-interpretation to medical diagnosis
	Davis et al.	• Joint histograms of image measurements (derived from the grey-level) at two points a fixed distance apart • General predicate relations between features allowed	• Generalization of above technique
Linear Filtering Models	Therrien	• Describes texture by a 2-d auto-regressive moving average (ARMA) model • Classifies texture on the basis of which model has the smallest prediction error	• Assumes texture can be synthesized by driving the ARMA model with white noise
Local Histograms	Wong & Shen	• Forms local histograms of image grey-levels and features • Computes a measure of texture similarity between histograms	• Similarity measures are invariant with respect to magnification and rotation of textures
Texture Energy/ Power Spectrum	Bajcsy	• Estimates power spectrum locally by Fourier transform • Uses energy computed in wedge/rings in the frequency domain as textural features	• Accuracy of Fourier components decreases as window size decreases
	Chen	• Uses the Lim & Malik 2-d maximum entropy spectral estimator	• Better resolution than DFT • Assumes AR model for texture
	Laws	• Uses "texture energy" passed by set of spatial filters as textural features	• Spatial-domain analog of above DFT approach
Structural	Tomita et. al.	• Groups textural elements (texels) into homogeneous regions • Computes properties of regions • Partitions feature space by thresholding histograms of texel properties	• Useful when textures are better described by their structural (as opposed to statistical) properties
	Fu	• Uses web and tree grammars to describe structural aspects of texture • Parses texel primitives using predefined grammars for each class • Grammars inferred through training or defined explicitly	• Handles the structural component well but has problems with noise and irregularities

test on the prediction errors from a bank of recursive filters, one for each texture.

D.3 LOCAL HISTOGRAMS

Wong and Shen (Ref. 101) represent texture using local histograms of image grey-level, and gradient-magnitude and gradient-angle. They develop similarity metrics for line frequency diagrams (i.e., histograms of linearly ordered measurements) and circular frequency diagrams (i.e., histograms of angular measurments). These metrics are computed over various image measures at several resolution levels, and are used to compute the similarity between different textures for clustering and classification.

D.4 POWER SPECTRUM/TEXTURE ENERGY

Early attempts to classify objects on the basis of texture used power spectrum measurements as textural features (Ref. 102). Given a finite number of samples, the resolution in the frequency domain may be traded off for the uncertainty (variance) in the spectral coefficients using the discrete Fourier transform (DFT). Small processing windows should be used to localize the textural measurements; thus, the accuracy of the DFT coefficients is poor. An alternate approach is to assume an underlying autoregressive process, and use a maximum entropy technique (Ref. 103) to compute the spectral coefficients as suggested by Chen (Ref. 104). Laws (Ref. 33) pursues an analogous approach in the spatial domain by using the outputs from a set of spatial filters as texture measurements. Here, the "texture energy" is computed in space, while in the former approaches it is computed over small regions (wedges and rings) in the frequency-domain.

D.5 STRUCTURAL

The above classes of techniques are examples of statistical representations for texture. Structural approaches (Ref. 105) typically involve first grouping regions in the image into texture elements (texels). Measurements such as intensity, area, shape, and directionality of texels are then computed. Texture classes may then be described in terms of these measurements to allow subsequent discrimination and classification.

Syntactic methods have also been used to build structural descriptions of texture. Fu (Ref. 106) has developed web and tree grammars to generate and recognize texture. While the syntactic approach handles the structural aspect of the texture quite well, the presence of noise and irregularity (both in the texels and in the repetition pattern) cause problems in the inference of texture grammars and in the parsing of textural patterns.

APPENDIX E
CLASSIFICATION

This appendix describes several major classes of classification techniques including: decision-theoretic and syntactic classifiers, and inference systems. A summary of representative techniques for classification is contained in Table E-1.

E.1 DECISION THEORETIC CLASSIFIERS

Decision theoretic classifiers represent objects statistically as groups or clusters in multi-dimensional space. Two basic approaches for constructing decision theoretic classifiers involve estimating (or assuming) the underlying probability densities for the object classes, and defining decision rules based on the class statistics (parametric approach) or computing the decision functions directly from a partition of the event space (non-parametric). The maximum likelihood classifier is an example of the former approach, where an event is assigned to the class having the highest a posteriori probability (the a priori probabilities are assumed to be equal). An example of a non-parametric statistical classifier is the Fischer linear discriminant function (Ref. 107) which projects multi-dimensional space onto the line which most effectively discriminates between (separates) selected classes.

TABLE E-1

SUMMARY OF CLASSIFICATION TECHNIQUES

GENERAL CLASS	EXAMPLE TECHNIQUE	SUMMARY	COMMENTS	APPLICATION AREAS
Decision Theoretic Classifiers	Maximum Likelihood Classifier	Assumes object class may be modeled by probability density function (e.g., Gaussian). Statistics used to determine decision rules.	Parametric technique	Multi-spectral and Textural Pattern Recognition
	Descriminant Function	Assumes no underlying density Partitions decision space into disjoint regions	Non-parametric technique	
Syntactic Classifiers	Context-Free String Parser	Objects are decomposed into primitives Input symbols (primitives) are parsed Objects are syntactically correct (i.e., recognized) if terminal state of parser is reached	Recognizes objects based on structure properties	Textural and Structural Pattern Recognition
Inference Systems	Production System	All knowledge is stored in "if-then" (production) rules Rules are evaluated by the *inference engine* to deduce new facts	Production rules contain all the knowledge required to determine object class	Expert Systems
	Probabilistic Relaxation Labelling	Labels and probabilities are associated with objects Probabilities are iteratively updated Labels with the highest probability are selected	Attempts to arrive at globally consistent interpretation of object	Vision Systems

E.2 SYNTACTIC CLASSIFIERS

Syntactic classifiers exploit the structure of an object for classification purposes. The structure may be the boundary of the object, or the repetitive pattern of object primitives (e.g., the texels used in structural analyses of texture). The pattern classifier is a finite state machine (Parser) which processes input symbols (object primitives) and outputs terminal symbol(s) which signify the recognition of pre-determined object classes. Syntactic techniques have been successfully applied to the recognition of objects on the basis of shape (contour) and to texture classification. They are generally sensitive to noise and irregularity in the data.

E.3 INFERENCE SYSTEMS

Inference systems deal with data in symbolic form; i.e., with the properties and relations between segments (edges and/or regions). In production systems, all domain-dependent knowledge is stored in a set of "if-then" production rules. The rules of inference (domain-independent) are embedded within the inference engine. The inference engine evaluates the production rules against all known facts in order to generate new facts, leading to the deduction or verification of hypotheses (possible object classes, for example).

Relaxation labelling algorithms do not perform classification per se; rather, they are useful in arriving at a consistent labelling or interpretation of an event (i.e., the collection of segments which comprise an object, the collection of objects which comprise the scene, etc.). In probabilistic relaxation labelling algorithms, labels and probabilities

for the labels are associated with each segment (area and/or edge). Semantic constraints are contained in the compatability matrix which is used to iteratively update the labels and their probabilities. As the interation approaches steady state, the labels having the highest probability (likelihood) are assigned to the segments.

APPENDIX F
ADAPTIVE LMS TECHNIQUE FOR MULTI-BAND ENHANCEMENT

A new approach to image enhancement is described in this appendix which utilizes an adaptive least mean square (LMS) error technique for computing optimal image estimates based on correlated reference images and an optional frequency domain replacement step for replacing known information into the estimate (Ref. 163). The technique can be used to smooth SEASAT SAR data or to spatially enhance Thematic Mapper (TM) thermal data. The technique relies on the assumption that at some resolution level, registered imagery data is highly correlated across bands or data sets. This assumption has previously been used in techniques for multi-band *Landsat* registration which have exploited the fact that "edge" information is highly correlated among all data sets (Ref. 154). The LMS technique utilizes reference bands to generate an optimal linear estimate of a desired band (or image) in an adaptive manner. This linear estimate can then be optionally combined with the original image to form an enhanced image.

The adaptive LMS approach is used to estimate SAR reflectance from SAR and TM IR data sets. SAR reflectivity is usually correlated with surface roughness, material type, and wetness, whereas IR imagery is indicative of "color" reflectivity. Although the SAR and IR data are uncorrelated on a global scale, the TM IR data set can be used on a local scale to estimate a smooth SAR reflectance map. The validity of this procedure relies on the fact that SAR imagery can be modeled by a smooth reflectance image plus a Rayleigh distribution of speckle noise.

In order to sharpen TM thermal data, the visible and IR imagery bands are used to estimate a high resolution thermal estimate using the adaptive LMS approach. The high spatial frequency information from the thermal estimate is then added to the original thermal data that is inherently low bandwidth using a frequency replacement method. This procedure forces the enhanced thermal data to be entirely consistent with the original thermal data set. In general, this approach can be used to interpolate (at an effective higher resolution) low resolution data when another highly correlated higher resolution data set is available.

In Section F.1 below, some factors are discussed motivating the multi-band enhancement procedure. Section F.2 then describes the adaptive multi-band LMS enhancement algorithm in detail. In addition, some experimental error results are tabulated regarding the choice of the adaptive window size and number of reference bands used. Finally some examples of processed SEASAT SAR and TM thermal data are given in Section F.3.

F.1 MOTIVATION

In utilizing multi-band/multi-source imagery, one is often faced with the problem of analyzing data that exhibit fundamental differences. For instance, TM thermal data is four times as coarse (spatially) as TM IR or visible data, and SAR data has high spatial resolution but contains high amounts of speckle noise compared to TM data (which is relatively smooth). This data incompatibility can adversely affect classification and other image processing procedures. For example, pixel-wise classification procedures that utilize TM thermal along with IR or visible data can err along object boundaries due to the coarser resolution of thermal data. The same classification procedures using SAR and TM data can yield very noisy results due to the

speckle noise content of SAR data. In the future, the role of multi-source/multi-band image processing will become increasingly important, and not only will it become necessary to combine new data sources but also to combine new and old data sources. Therefore, a multi-band LMS or similar technique is deemed necessary to support preprocessing of multi-source/multi-band data.

In order to use these diverse types of data sources in an optimal manner, one can utilize specific information from the contributing data sources in order to improve a particular characteristic of any one source. This characteristic can typically be spatial resolution or noise level. One can accomplish this by using inherent local correlation between data sources. In order to properly exploit this correlation assumption, an adaptive multi-band approach was developed. At the coarsest resolution the adaptive window is the size of the image and, as a consequence, the data correlation between the desired image and the reference images is expected to be low. As the window size decreases, the correlation increases. In the next section the details of the algorithm are discussed.

F.2 APPROACH

Given the assumption that multi-source images are correlated at some resolution level, one can use the contributing images as basis functions (at the local level) to predict or estimate any of the individual images. In practice, the images that have high resolution and low noise will be used to estimate the images that have low resolution (TM thermal data) or high noise (SEASAT SAR data). The multi-band enhancement procedure entails two steps. The first step is to compute the optimal estimation of the degraded image using an adaptive LMS predictor. (This first step can be used alone to estimate SAR reflectance images.) The second step is to filter the estimate and combine

it with the degraded image to form an enhanced result. The TM thermal data is subsequently processed in this manner. The two processing steps are described in detail in the next subsections.

F.2.1 Adaptive LMS Multi-Band Predictor

Using a linear prediction approach, the desired signal (image) estimate is formed by a weighted linear combination of reference images, where the weights change adaptively over the entire image. This construct is analogous to adaptive LMS filters for one-dimensional signals. A synchronized two-dimensional window slides over the degraded image (to be estimated) and the reference images. The output of the windows can be treated as vectors of data. The coefficients b_i form the estimate $\hat{y}$ (for the center pixel of the window) while being continually adjusted to minimize the error vector $\underline{e}$. For a given location of the sliding prediction window, one can represent the (to be estimated) image $\underline{y}$ as

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_w \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdot & \cdot & \cdot & X_{p1} \\ 1 & x_{12} & x_{22} & \cdot & \cdot & \cdot & X_{p2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{1w} & x_{2w} & \cdot & \cdot & \cdot & X_{pw} \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_p \end{bmatrix} + \begin{bmatrix} e_o \\ e_1 \\ \cdot \\ \cdot \\ \cdot \\ e_p \end{bmatrix} \qquad \text{(F-1)}$$

or in vector matrix form as

$$\underline{y} = X\underline{b} + \underline{e} \qquad \text{(F-2)}$$

where the matrix X contains the concatenated vectors of reference data as well as a constant vector. One can also write (F-2) as

$$\underline{y} = \hat{\underline{y}} + \underline{e} \qquad \text{(F-3)}$$

where $\hat{\underline{y}}$ is the optimal estimate for $\underline{y}$. The number of reference images is equal to p, the number of data values within each window is equal to w, and $w \geq p$.

The LMS solution for the predictor $\hat{\underline{y}}$ is computed by solving for the set of coefficients $\underline{b}$ which minimize the following error norm:

$$||\underline{e}|| = (\underline{y} - X\underline{b})^T(\underline{y} - X\underline{b}) \qquad \text{(F-4)}$$

Minimizing the quantity in (F-4) is equivalent to solving the set of normal equations,

$$(X^TX)\underline{b} = X^T\underline{y} \qquad \text{(F-5)}$$

for the coefficient vector $\underline{b}$. In one-dimensional signal linear prediction, equation (F-5) can be efficiently solved using a recursive approach. Since the one-dimensional window update involves adding and subtracting only one data value, the matrix XX^T can be updated by shifting and adding one column and row of data. For the two-dimensional case, however, the two-dimensional window update involves adding an entire column of new image data. Therefore, the updating procedure for the matrix XX^T for the two-dimensional window becomes more complicated. It turns out that the order of complexity for the two-dimensional update is comparable to the direct solution of (F-5) (Ref. 117). As a consequence, our implementation of the two-dimensional linear prediction involved the direct solution:

$$\underline{b} = (X^TX)^{-1}(X^T\underline{y}) \ . \qquad \text{(F-6)}$$

The execution of this procedure yields an optimal estimate, $\hat{\underline{y}}(n,m)$, which has been computed adaptively over the domain of the two-dimensional sliding window. For each location (n,m)

of the sliding window, one corresponding value of $\hat{y}(n,m)$ was computed using the current values of the linear prediction coefficients.

F.2.2 Spatial Frequency Replacement

The frequency replacement step is a nonlinear procedure to combine data in the spatial frequency domain. For a given spatial frequency range, the replacement algorithm allows one to replace the phase or magnitude (or both) of a signal with the phase and/or magnitude of the computed optimal estimate of the signal. In order to enhance the spatial resolution of data, the replacement procedure reduces to lowpass filtering the degraded data, highpass filtering the optimal estimator, and adding the filtered results; i.e.,

$$\hat{Y}(k,l) = H_{\ell p}Y(k,l) + H_{hp}\hat{Y}(k,l) \; . \tag{F-7}$$

where the cutoff frequency of $H_{\ell p}$ is chosen as low as possible such that

$$Y(k,l) = H_{\ell p}Y(k,l) \tag{F-8}$$

The highpass filter is the complementary filter given by

$$H_{hp}(k,l) = 1 - H_{\ell p}(k,l) \tag{F-9}$$

One interpretation of (F-7) is that the low spatial frequency information is preserved (the truth data is not distorted) and it is supplemented with high frequency information computed from an optimal estimate.

F.2.3 Window Size and Reference Bands Considerations

An error analysis was performed for the estimation of band 7 (far IR band) of TM data by the other TM bands using varying window sizes and numbers of reference bands. The scene chosen was an agricultural area in Kansas. An rms error (root of the error norm in (F-4)) was calculated for each window location and then averaged over all locations. These RMS error figures are tabulated in Table F.2-1. From this table one observes that the error decreases rapidly with decreasing window size. Secondly, increasing the number of reference bands will also decrease the estimation error. For large windows, however, increasing the number of reference bands will not significantly reduce the error.

TABLE F.2-1

AVERAGE RMS ESTIMATION ERROR AS A FUNCTION OF NUMBER OF BANDS AND WINDOW SIZE

BANDS	WINDOW SIZE				
	63 × 63	31 × 31	15 × 15	7 × 7	3 × 3
1	17.58	14.36	12.03	8.25	4.43
2	17.29	13.97	11.74	7.83	3.71
4	12.51	9.98	8.35	5.55	2.08
6	10.32	7.89	6.44	4.37	1.14

F.3 IMAGERY EXAMPLES

An initial version of the adaptive LMS technique has been implemented as illustrated in Fig. F.3-1. The processed examples in this section include estimating SAR reflectance imagery and spatially sharpening LANDSAT TM middle IR and thermal IR imagery.

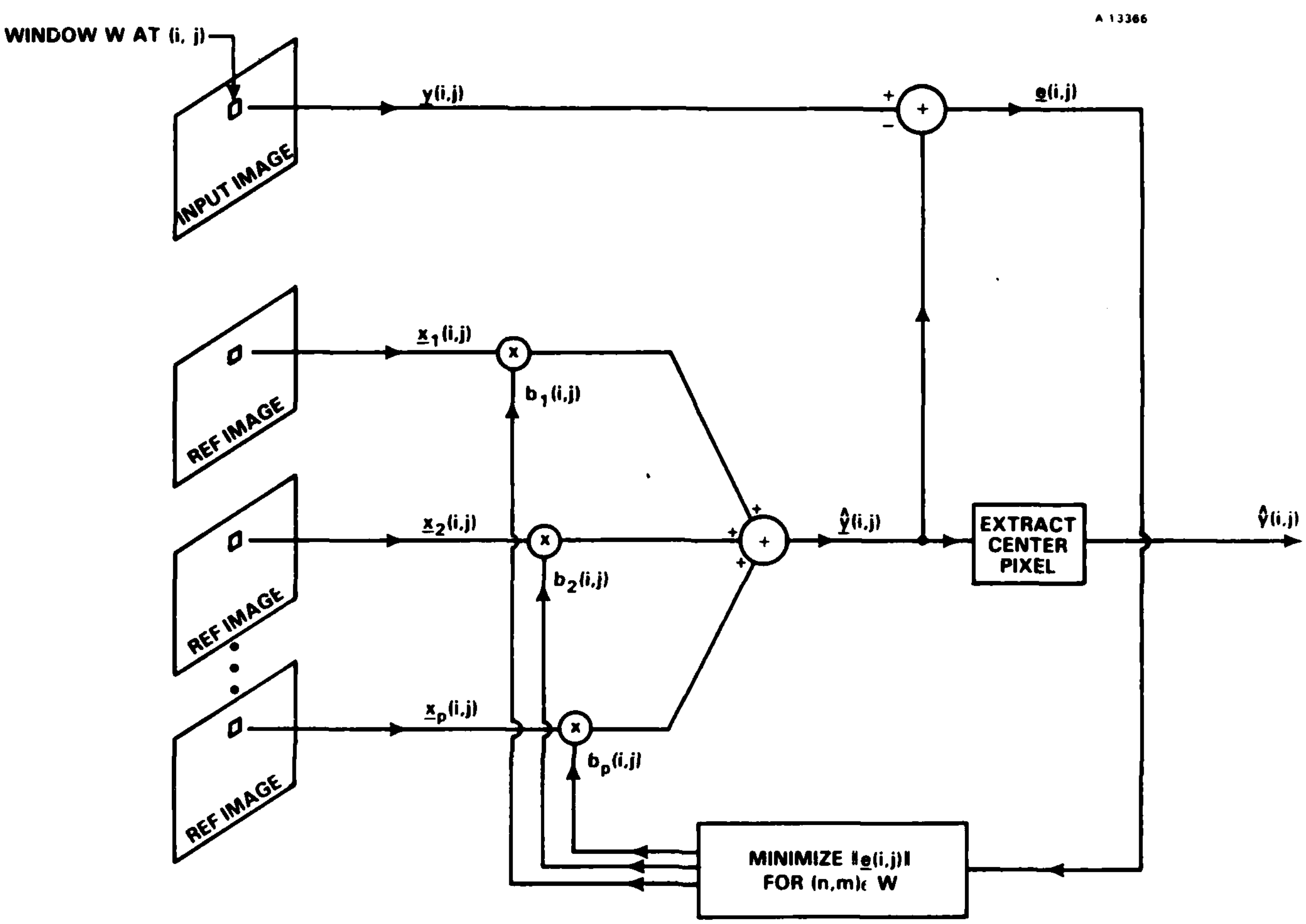


Figure F.3-1 Block Diagram of Adaptive LMS Technique

As a test of the procedure, we first degraded a TM middle IR image (band 7) from 30 m to 120 m resolution and then restored it to 30 m resolution. An original TM band 7 image is presented in Fig. F.3-2. Thematic Mapper bands 1-5 are used as reference images and one of these, band 4, is presented in Fig. F.3-3. A simulated TM band 7 at 120 m resolution (Fig. F.3-4) was generated by blurring with a 4 × 4 averaging window and then subsampling. The reference bands were also pre-blurred and subsampled before the LMS method was applied. Using the set of interpolated coefficients on the full resolution reference bands produced the restored TM band 7 in Fig. F.3-5. A comparison

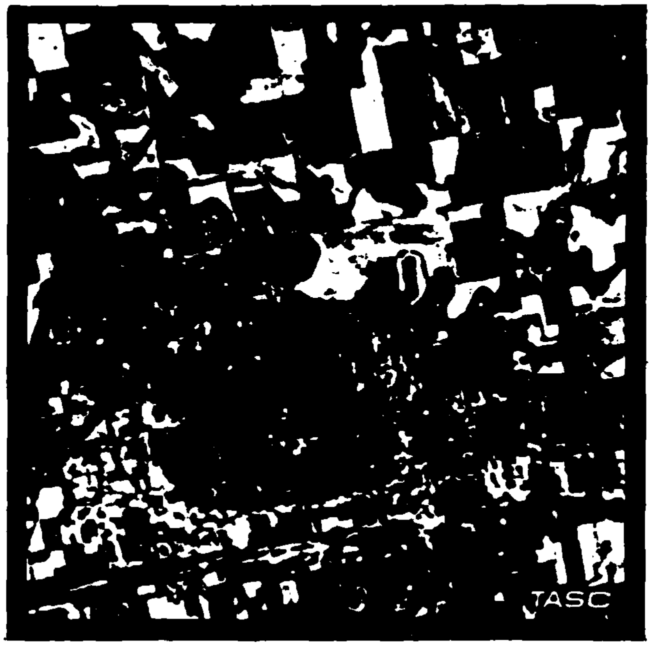


Figure F.3-2 Original Thematic Mapper Band 7 (Far IR)

Figure F.3-3 TM band 4 (Near IR)

Figure F.3-4 TM band 7 at 1/4 Resolution

Figure F.3-5 TM band 7 Restored

of the mean-square error (original-blurred) before enhancement (Fig. F.3-6) and the error (original-restored) after enhancement (Fig. F.3-7) shows the reduction of error especially along object borders. Much of the residual error is due to the ill-conditioning of the matrix in (F-6) for regions that are relatively flat and highly correlated. This example provides a reasonable justification for the application of the LMS technique.

The LMS procedure was also applied to TM thermal IR (band 6) data. The original thermal data is presented in Fig. F.3-8. Note the similarity of the overall blurring with the simulated degradation of Fig. F.3-3. Bands 1-5,7 were used as reference bands and were pre-blurred before applying the LMS technique. After the LMS procedure was applied, the Fourier replacement step was used to make the final enhanced thermal imagery consistent with the original data. The final enhanced image is presented in Fig. F.3-9. In practice, the LMS procedure alone produces an estimate whose spectrum closely approximates the original data so that the Fourier replacement step may not be necessary.

To provide an example of how the LMS technique can be applied to multi-source imagery, the technique was used to smooth SEASAT SAR imagery. An original SEASAT SAR image is presented in Fig. F.3-10. Three co-registered TM images are used as reference images (Fig. F.3-11). The window size was 15 × 15. In order to estimate the average reflectance, the high specular returns are first detected (Fig. F.3-12) and filtered out. Then the LMS technique is applied using the TM images to optimally estimate the SAR reflectance. The final filtered SAR image, with the specular points overlaid, is presented in Fig. F.3-13. The result shows a significant reduction in the speckle level, without noticeable edge degradation.

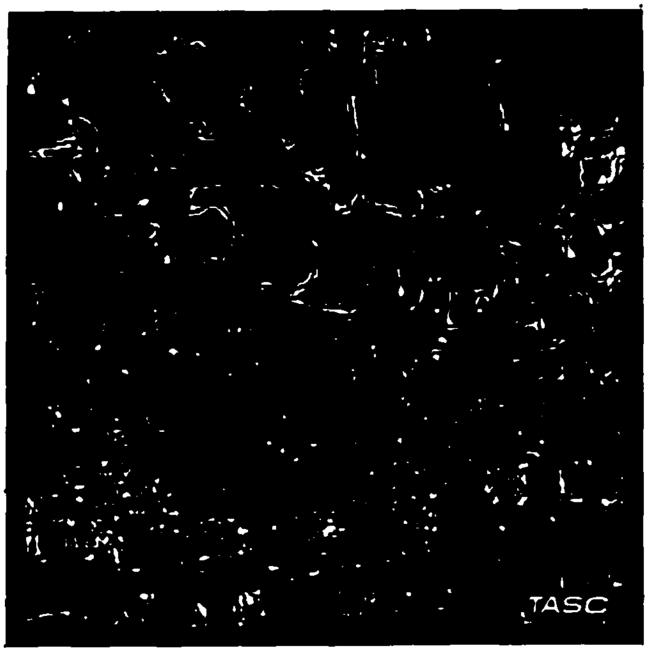


Figure F.3-6 Mean-Square Error Before Enhancement

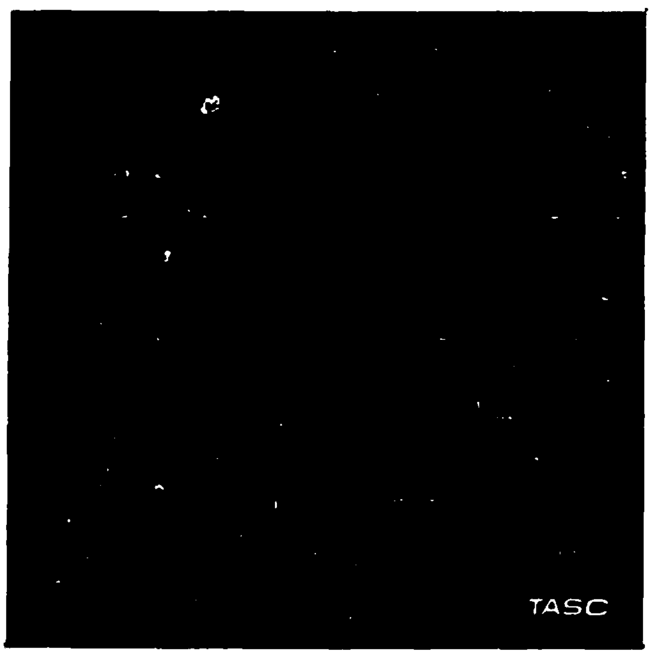


Figure F.3-7 Mean-Square Error After Enhancement

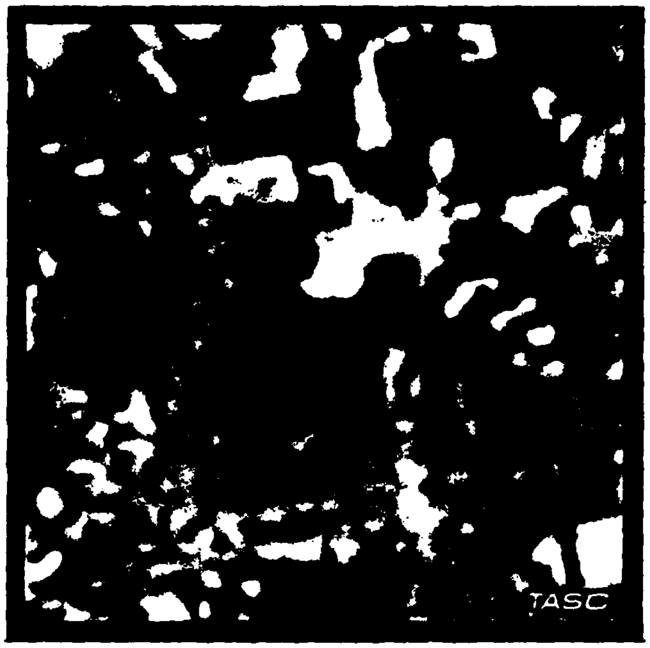


Figure F.3-8   Original TM band 6 (Thermal)

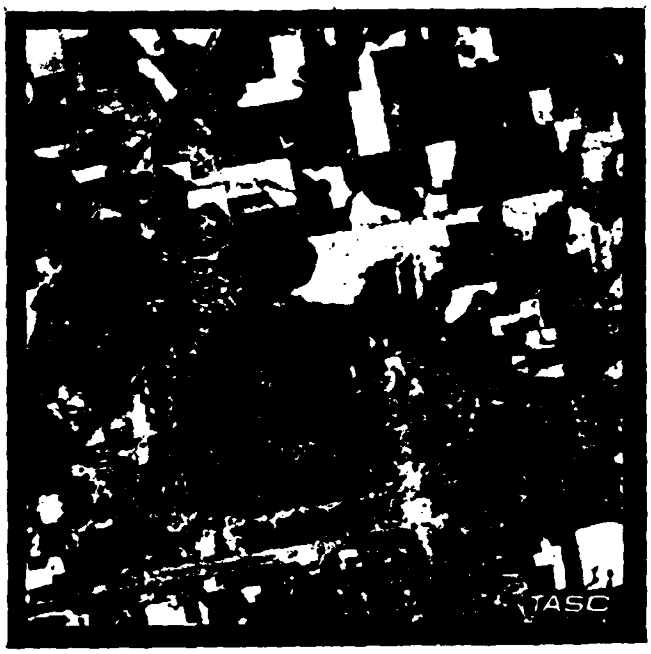


Figure F.3-9   Enhanced TM band 6 (Thermal)

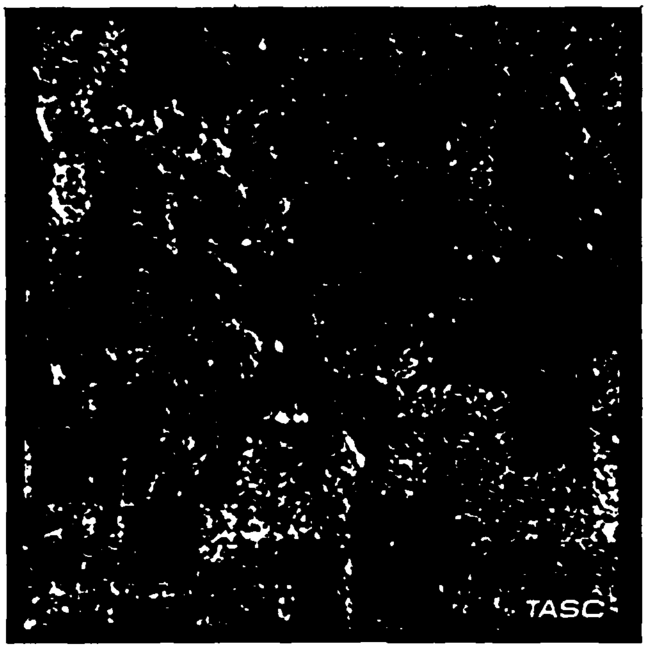


Figure F.3-10 SEASAT SAR Image

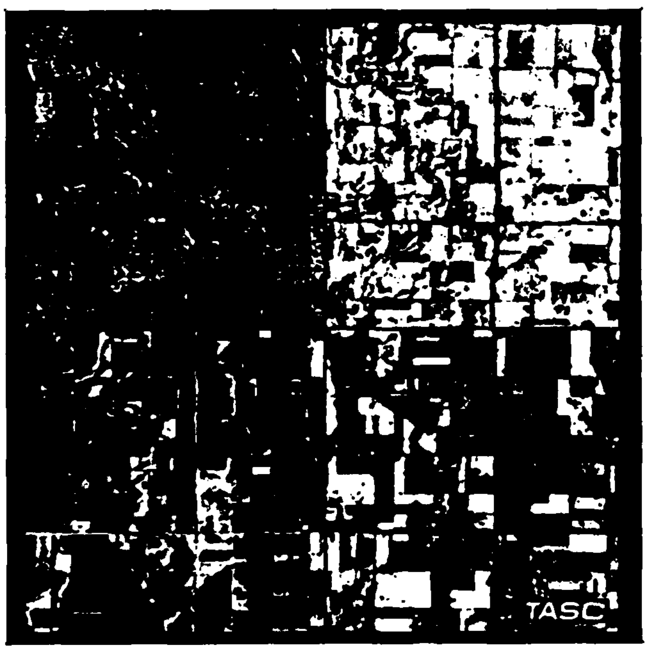


Figure F.3-11 SAR and Three Registered TM Images

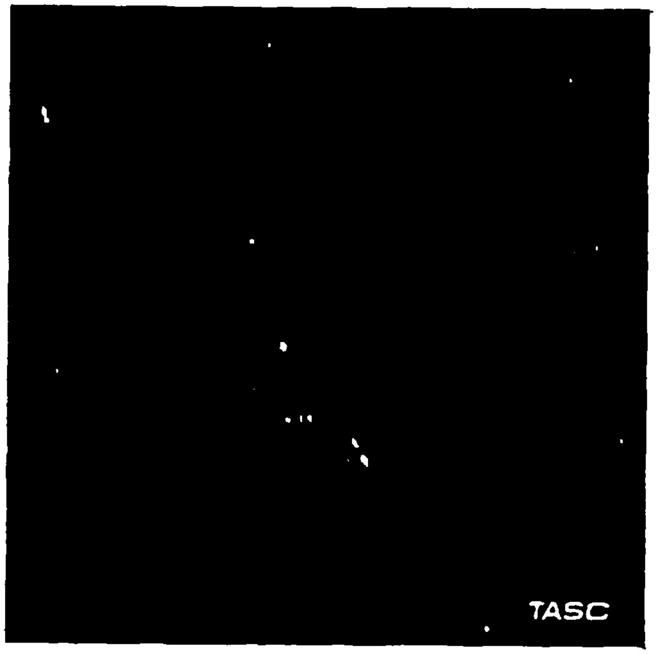


Figure F.3-12 Specular Reflections

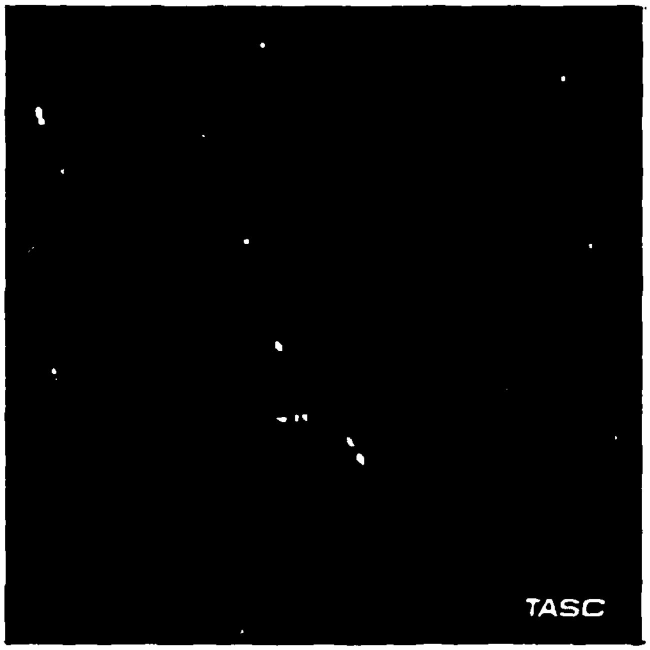


Figure F.3-13 Adaptively Filtered With Specular Parts Overlaid

F.4 SUMMARY

A new approach to image enhancement was described which utilizes an adaptive least mean square error technique for computing optimal image estimates based on correlated reference images and an optional frequency domain replacement step for replacing known truth information into the estimate. The technique was used to smooth SEASAT SAR data and to spatially enhance Thematic Mapper (TM) thermal data. The technique is useful for producing co-registered MS/MS data sets with the same effective spatial resolution.

APPENDIX G
KNOWLEDGE-BASED MULTI-SPECTRAL IMAGE CLASSIFICATION

A knowledge-based approach for classifying surface materials in multi-spectral imagery is outlined in this appendix. Surface material classes are defined heuristically using rules which describe the typical appearance of the material under specified conditions in terms of relative image measures. The knowledge-based approach allows expert knowledge of the domain to be used to develop classification rules directly without training. To illustrate the approach, a LANDSAT TM image is classified into general land-cover categories.

G.1 MOTIVATION

Land-use monitoring has been one of the many applications which has motivated the development of analysis technologies for earth remote sensing. In most classification processes to date, land cover or surface material classes have been represented statistically (e.g., in terms of the means and covariances of spectral bands and transformations of spectral bands in prototypical regions in a training data set). This statistical description is then used to define decision boundaries in feature space which are subsequently used to classify test data set(s). In attempting to use the training statistics computed in one image to classify another, signal variability becomes a problem. For example, one image may be hazier than the other, or images taken at different times may be different due to changes in illumination or biomass. Thus,

while there is generally sufficient information at the signal level for discrimination, the information typically is not sufficient for classification over a wide range of conditions. In short, a unique and invariant signature for each surface material or land cover class does not exist.

Anticipating the increased resolution of future multi-spectral sensors, increased coverage requirements, and reduced interpretation timelines, an alternate approach to statistical (signal-based) classification is needed to increase the degree of automation possible in a multi-spectral MC&G feature extraction system. The approach described below uses rules to define the typical or expected appearance of surface materials in terms of <u>relative</u> image measures.

G.2 APPROACH

An image analyst (IA) familiar with a particular type of sensor is generally able to recognize surface materials over a wide range of scene conditions. IAs are able to interpret imagery under variable conditions because they know the kind of scene they are looking at (hence the types of objects and materials to expect in the scene), and are able to reason about the appearance of various materials and structures in the image not only in the visible, but in the infrared and microwave regions as well. Since humans tend to describe things in relative terms (e.g., wet fields are darker than dry fields in the visible and infrared) rather than absolutely, it seems appropriate to develop a representation which is based on relative image measures.

Two kinds of relative image measures are currently being investigated for the purpose of characterizing surface

materials over a wider range of scene conditions than is now possible with absolute signal representations. These relative image measures are computed by analyzing the images pixel-by-pixel across wavelength (spectral signature analysis), or at a particular wavelength or spectral band across intensity (histogram analysis).

Spectral signature analysis at a particular point in the image gives a view of the intensities across all bands. Thus, one can speak of trends in the signature (e.g., the spectral response peaking at a particular wavelength), and use this qualitative description to classify certain materials. As an example, a simple signature analysis of LANDSAT TM bands 4, 5, and 7 permits vegetation and soils to be easily separated. The cell structure of vegetation (crops, trees, grass) causes most of the incident radiation in band 4 to be either reflected or transmitted, and much of the radiation in band 5 to be absorbed due to water in the cell structure. Soils, on the other hand, tend to reflect increasing amounts of radiation as one progresses into the far infrared.

The histogram analysis of a given band summarizes the relative frequency of intensities in that band over the entire image. If one knows something about the contents of the scene, it is possible to relate modes in the histograms (i.e., clusters in feature space) to instances of particular materials in the image. For example, since the reflectivity of water, in the infrared is quite low, if the scene contains water then the darkest regions in the infrared are likely to be water. A simple approach, then, for extracting water is to identify the darkest modes in the infrared intensity histogram, and label pixels having IR intensities within a prescribed range as water. An example of this procedure is illustrated in Appendix H.

It is apparent that the use of relative image measures mitigates such effects as uniform haze and variations in the sensor gain between images. It is not clear at this time to what extent the ability to distinguish similar SMCs is reduced by using relative image measures.

G.3 LANDSAT-TM CLASSIFICATION EXAMPLE

An initial version of an expert system has been implemented to demonstrate the rule-based approach. A forward-chaining inference engine applies the rules described below to the image on a pixel-by-pixel basis. Histogram analysis is performed initially to compute thresholds for mode selection. In order to determine if a pixel belongs to a particular mode in a spectral band, the value of the pixel is simply compared to upper and lower thresholds. Similarly, relative responses between bands are computed by comparing the values of those bands.

LANDSAT TM imagery over Lawrence, Kansas was used in our demonstration. Bands, 2, 4, and 7 from such a TM image are shown in Fig. G.3-1 in blue, green, and red, respectively. USGS topographic maps and aircraft photos were used to infer ground truth for developing rules and testing the classifier.

In our demonstration, the scene was first partitioned into three major classes (water, vegetation, and soil-like materials) by applying the following rules, in order, to the image:

Figure G.3-1 LANDSAT-TM Image Over Lawrence, Kansas

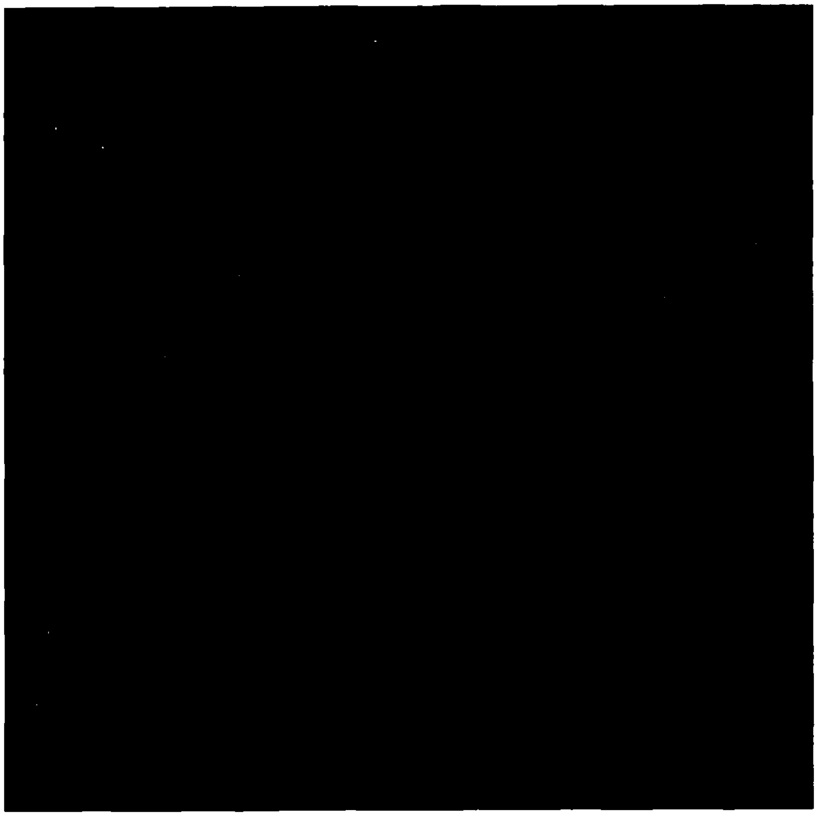


Figure G.3-2 Classification Image

If: (Band-4-relative-intensity = DARK)
Then: (assert water)

If: (Band-4-intensity>Band-3-intensity) and (Band-4-intensity>Band-5-intensity)
Then: (assert vegetation)

If: (Band-5-intensity>Band-4-intensity)
Then: (assert soil-like)

Other rules were developed to decompose these major classes into sub-classes; i.e., soil-like materials into plowed fields and concrete/silt, and vegetation into crops. Plowed fields have the general spectral characteristics of soils (i.e., increasing response in the IR), but tend to be darker than most soils due to higher moisture and organic content. On the other hand, concrete, which also has the characteristic signature of soils in the IR, is bright relative to other soil-like materials in the visible. In this particular scene, crops gave rise to the highest response in band 4, although the usefulness of this discriminant will depend on the period of the growth phase. (This being one case in which collateral information would be required.) The classification image is shown in Fig. G.3-2 with water in blue, vegetation in green (crops are bright green, other vegetation such as trees and grass are dark green), and soil-like materials in yellow (concrete/silt), red (plowed fields), and brown (uncultivated areas). The resultant classifications compared favorably with ground truth inferred from USGS topographic maps.

G.5 SUMMARY

A preliminary version of a multi-spectral image classification expert system was described and assessed. The

rule-based approach represents an alternate approach to decision-theoretic statistical classification and has the potential to increase the degree of automation possible in multi-spectral DMA feature extraction systems. While initial results were promising, additional research is required.

APPENDIX H
IMAGE SEGMENTATION VIA HISTOGRAM ANALYSIS

A new method for analyzing histograms for image segmentation is described in this appendix. The technique does not rely on the presence of peaks in the histogram for splitting, nor does it require that the number of modes be known in advance. Modes in the histogram are identified by first smoothing (convolving) the histogram with a Gaussian to remove spurious peaks and then marking the location of zero-crossings in the first and second derivatives of the smoothed histogram. Zero-crossings in the first derivative are extrema (peaks and valleys) in the smoothed histogram, and zero-crossings in the second derivative are turning points (points of inflection). Modes are detected by locating particular sequences of zero-crossings in the histogram. The smoothed histogram is approximated by a Gaussian mixture. Initial estimates of the mixture parameters (i.e., the mean, variance, and relative frequency of each component density) provided by the zero-crossing analysis are subsequently refined using an iterative maximum likelihood estimator. The technique is used within the knowledge-based surface material classifier described in Appendix G for finding regions that are relatively dark/bright in particular spectral bands.

H.1 MOTIVATION

The goal of image segmentation in general is to partition the image into physically meaningful units; i.e., regions having the same surface orientation, depth, or composition.

In systems which use color or multi-spectral imagery, the specific goal is to partition the image into spectrally homogeneous regions. An assumption underlying feature-based (as opposed to image-based or region growing) techniques is that each type of region in the image gives rise to an associated distribution in feature space. In reality, distributions from different kinds of regions often overlap (in one dimension, the histogram may not exhibit well-defined peaks). Thus, techniques which rely on the presence and/or formation of well-defined peaks in histograms may fail to identify clusters in the data.

H.2 APPROACH

An analysis of the histogram of a selected spectral band is performed at a particular scale or resolution by convolving the histogram with a Gaussian, and marking the locations and signs of the zero-crossings in the first and second derivatives. The histogram is approximated at the selected scale by a Gaussian mixture. The number of component densities and initial estimates for their mean, standard deviation, and relative frequency are provided by the histogram analysis. These estimates are subsequently refined using an iterative maximum likelihood estimation technique. The spectral band is segmented by thresholding the histogram so as to minimize the average probability of mis-labeling a segment.

H.2.1 Analysis of Gaussian Densities

The one-dimensional Gaussian density

$$g(x) = \frac{1}{2\pi\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad \text{(H-1)}$$

is fully characterized by its mean μ and standard deviation σ. The first and second derivatives of the Gaussian are given by

$$\frac{\partial g}{\partial x} = \frac{(\mu - x)}{2\pi\sigma^3} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{H-2}$$

and

$$\frac{\partial^2 g}{\partial x^2} = \left[\frac{(\mu - x)^2}{\sigma^2}\right] - 1 \quad \frac{1}{2\pi\sigma^3} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{H-3}$$

The first derivative has a zero-crossing (i.e., the sign of the slope changes from positive to negative) at the peak (mean) of the Gaussian. Zero-crossings in the second derivative are points of inflection and occur at $\mu \pm \sigma$.

Now consider a mixture of two Gaussian densities:

$$f(x) = \frac{p_1}{2\pi\sigma_1} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}} + \frac{p_2}{2\pi\sigma_2} e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}} \tag{H-4}$$

where $p_2 = 1 - p_1$. A mixture having parameters $\mu_1 = 130$, $\mu_2 = 160$, $p_1 = 0.9$, $p_2 = 0.1$, and $\sigma_1 = \sigma_2 = \sigma = 5$ is shown along with its first and second derivatives in Fig. H.2-1. The component densities are sufficiently far apart for two distinct peaks to form. The three zero-crossings of the first derivative correspond to two peaks and the one valley. In the second derivative, two pairs of two zero-crossings occur on either side of the peaks. If the component densities are far

enough apart, the peaks will be close and the means and the distance between turning points on either side of a peak will be approximately equal to twice the standard deviation.

In Fig. H.2-2 the standard deviations are increased to $\sigma = 10$ causing the smaller of the two peaks and the valley to disappear into the turning point between them. Only a single peak remains (one zero-crossing in Fig. H.2-2b) in the histogram. The presence of the second, smaller mode is still apparent in the second derivative (Fig. H.2-2c). In Fig. H.2-3, the standard deviations are further increased to $\sigma = 15$ causing the larger mode to completely dominate and obliterate the smaller mode. One zero-crossing remains in the first derivative and two zero-crossings remain in the second derivative.

The location of peaks, valleys, and turning points are plotted in Fig. H.2-4 for $\sigma = 15$, 10, and 5. A single mode is observed when $\sigma = 15$ (turning point → peak → turning point). Two modes are seen at $\sigma = 5$. It is obvious that the ability to resolve the two modes is dependent on the variance of the component distributions. However, by examining the second derivative, the presence of the second mode is apparent at $\sigma = 10$, even though there is no valley.

H.2.2 Fingerprints of Histograms

The plots in Fig. H.2-4 of the location of zero-crossings in Figs. H.2-1 through H.2-3 have the same structure as slices from the fingerprint of a one-dimensional signal (Ref. 155). Fingerprints depict trajectories of zero-crossings in the second derivative of a signal convolved with a Gaussian in scale-space; i.e., as the scale (standard deviation of the Gaussian) changes. Extrema of the n-th derivative are given in the zero-crossings in the (n+1)-st derivative. Thus, peaks

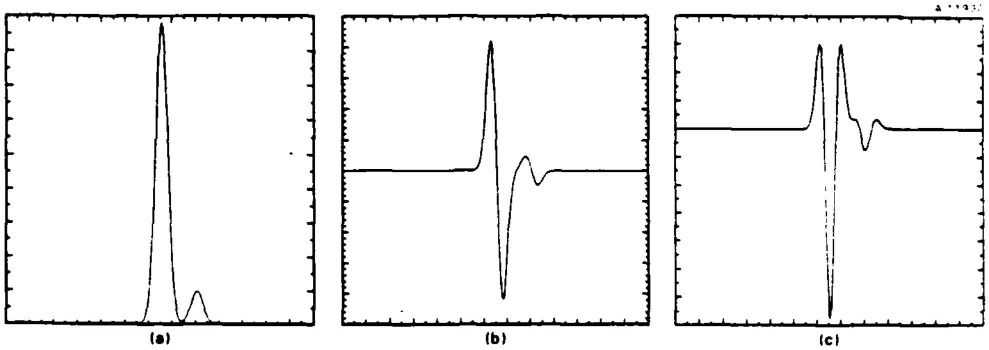


Figure H.2-1 Gaussian Mixture and Derivatives (Two Distinct Peaks)

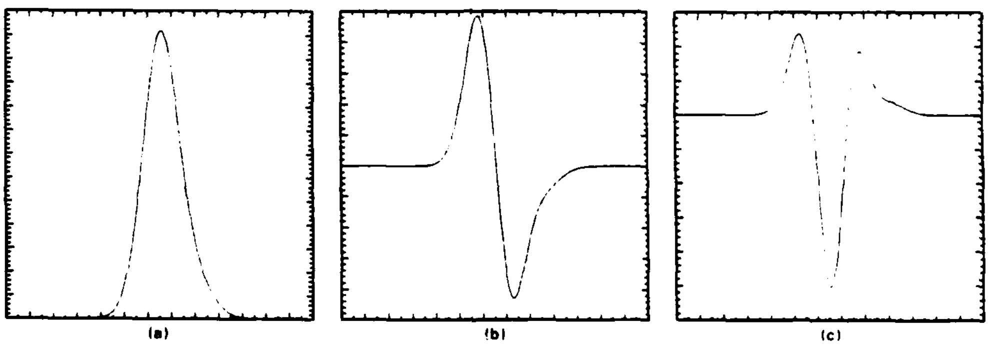


Figure H.2-2 Gaussian Mixture and Derivatives (Partially Buried Peak)

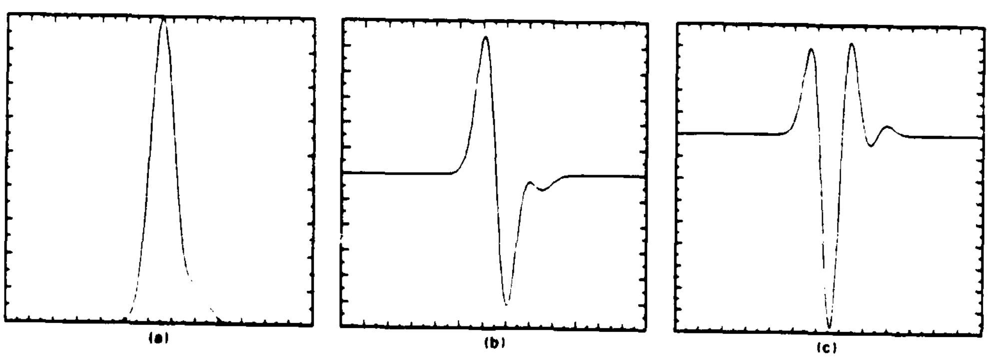


Figure H.2-3 Gaussian Mixture and Derivatives (One Peak)

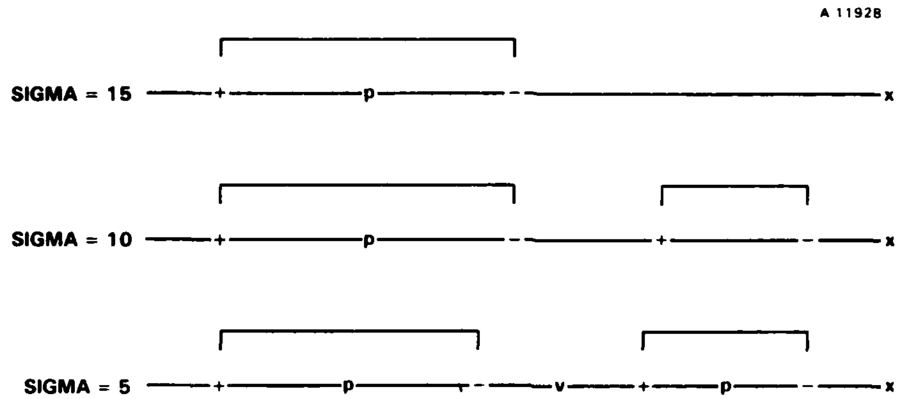


Figure H.2-4 Zero-Crossings in Previous Three Mixtures ("p" - peak, "v" - valley, "+" - plus-to-minus signal change in second derivative, "-" - minus-to-plus signal change in second derivative)

and valleys correspond to zero-crossings (plus-to-minus and minus-to-plus sign changes) in the first derivative. Turning points are zero-crossings in the second derivative.

Band 4 from the Landsat TM over Lawrence, Kansas and its histogram are shown in Fig. H.2-5. The fingerprint of the histogram is shown in Fig. H.2-6. The similarity in form between Figs. H.2-4 and H.2-7 has a simple explanation: a mixture of Gaussian densities and the convolution of a signal with a Gaussian are both weighted sums of Gaussians. A conclusion is that a mode in the mixture gives rise to a pair of turning points of opposite sign.

The above suggests a method for extracting the modes in a histogram at a particular scale:

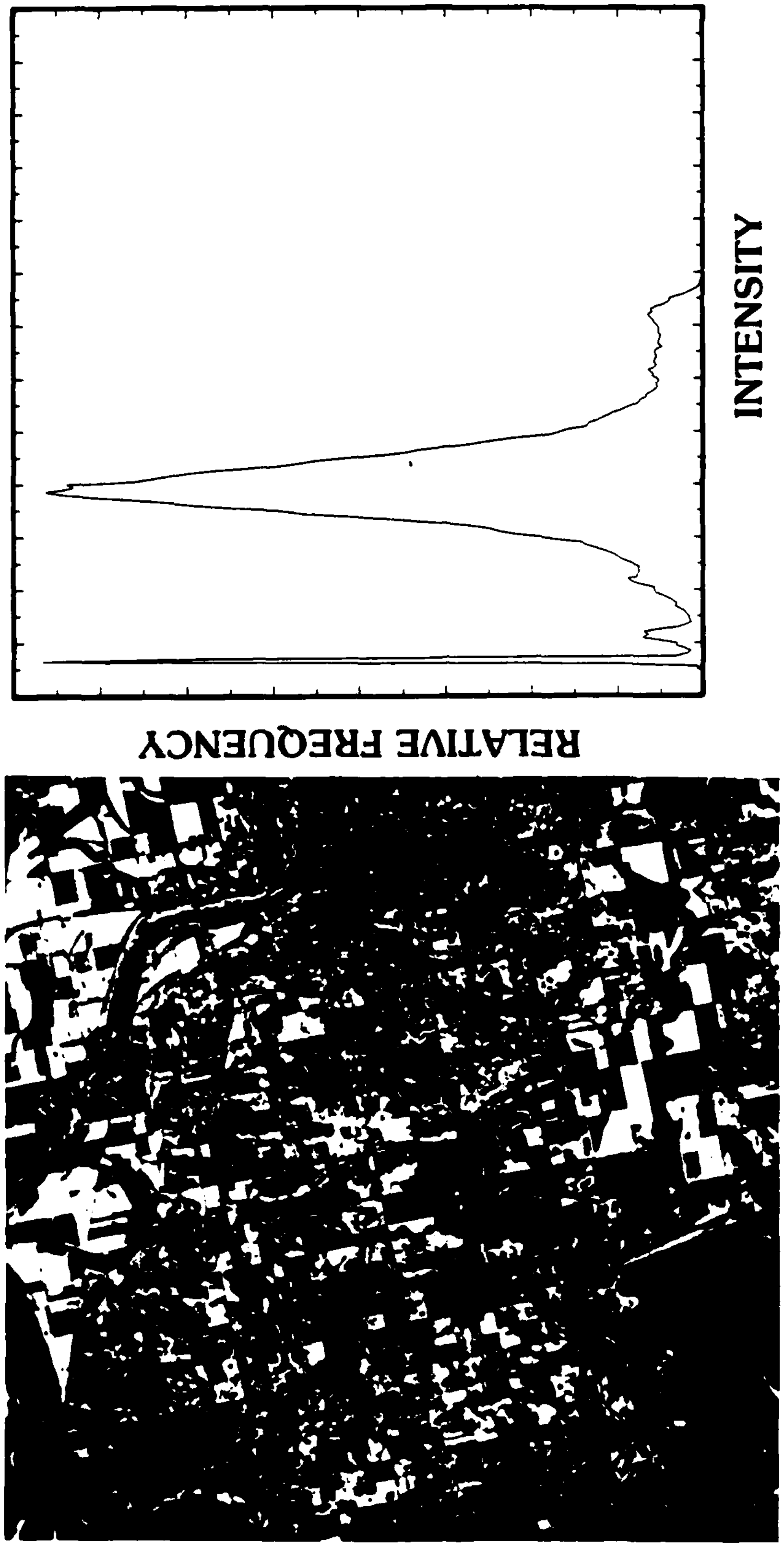


Figure H.2-5   Landsat TM Band 4 Over Lawrence, Kansas and Histogram

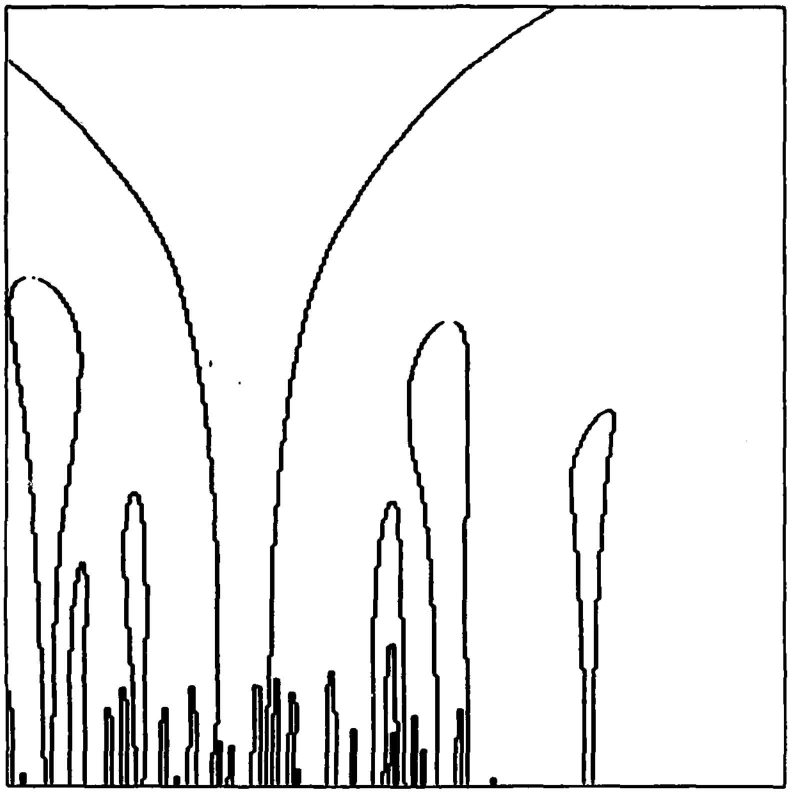


Figure H.2-6 Fingerprint of Previous Histogram

A 11927

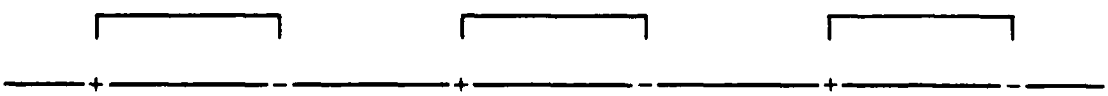


Figure H.2-7 A Slice From Previous Fingerprint
(Peaks and Valleys are not shown)

- Smooth the histogram by convolving it with a Gaussian
- Locate turning points in the smoothed histogram
- Extract modes by parsing the zero-crossings left to right (i.e., in the direction of increasing x) and grouping adjacent pairs of turning points (of opposite sign).

While a large amount of smoothing is desirable to reduce sampling noise (i.e., the jagged appearance of histograms resulting from a small statistical sample), as σ increases the ability to resolve closely-spaced modes in the data is reduced. (In general, modes separated by less than 2σ will not be resolvable.)

H.2.3 Estimating Parameters of Gaussian Mixtures

Having determined the number of modes at a particular scale, a model for the distribution of image intensities in each mode must be selected. In some domains, certain densities may be preferred (e.g., in SAR imagery, the radar backscatter from homogeneous regions is often characterized by a Rayleigh or log-normal density). It is assumed here that the histogram may be approximated by a mixture of Gaussian densities. What remains then is the determination of the parameters of the model. The parameters Θ of the mixture which "best fits" the observed data $\{x_i\}$ are the values for the relative frequency $p(\omega_q)$, mean μ_q, and standard deviation σ_q of each component Gaussian which maximizes the joint probability of observing the $\{x_i\}$ for i = 1,2,..., N where N is the number of samples:

$$p(X\,|\Theta) = \prod_{i=1}^{N} p(X_i\,|\Theta) \qquad \text{(H-5)}$$

If the number of classes Q is known and initial estimates are available, a solution to the maximum likelihood (ML) equations can be obtained iteratively (Ref. 156) according to

$$\hat{p}(\omega_q) = \frac{1}{N} \sum_{k=0}^{k-1} \hat{p}(\omega_q \mid x_k)\, f(x_k) \tag{H-6}$$

$$\hat{\mu}_q = \frac{\sum_{k=0}^{k-1} \hat{p}(\omega_q \mid x_k)\, f(x_k)\, x_k}{\sum_{k=0}^{k-1} \hat{p}(\omega_q \mid x_k)\, f(x_k)} \tag{H-7}$$

$$\hat{\sigma}_p = \frac{\sum_{k=0}^{k-1} \hat{p}(\omega_q \mid x_k)\, f(x_k)(x_k - \mu_q)^2}{\sum_{k=0}^{k-1} \hat{p}(\omega_q \mid x_k)\, f(x_k)} \tag{H-8}$$

where

$$\hat{p}(\omega_q \mid x_k) = \frac{p(x_k \mid \omega_q)\, \hat{p}(\omega_q)}{\sum_{q=0}^{Q-1} p(x_k \mid \omega_q)\, \hat{p}(\omega_q)} \tag{H-9}$$

is the <u>a posteriori</u> probability that x_k is from the q-th or distribution. Since the iteration will yield local maxima only, initial estimates close to the desired (globally optimal)

solution are required. Fortunately, the zero-crossing analysis computes the number of modes and can provide initial estimates for the mixture parameters. As initial (rough) estimates for the parameters of the component Gaussians, the location of the peak or, if a peak does not exist, the point halfway between turning points is used as an estimate of the mean, half the distance between turning points as an estimate of the standard deviation and the fraction of the total area of the histogram between the outer turning points (if a valley is present then out to the valley) as an estimate of the relative frequency are used.

Segmentation into disjoint regions is performed using a minimum probability of mis-classification criterion. The label corresponding to the class ω_q having the greatest <u>a posteriori</u> probability is assigned to each pixel. The decision criterion is to pick class p if

$$p(\omega_p | x_k) > p(\omega_q | x_k) \qquad \text{(H-10)}$$

for all $q \neq p$. If $p(\omega_p | \omega_q)$ is the probability of assigning a pixel to class ω_p given it really belongs to class ω_q, then the probability of error is

$$P_{error} = \sum_{q=0}^{(Q-1)} \sum_{p \neq q} p(\omega_p | \omega_q)\, p(\omega_q) \qquad \text{(H-11)}$$

where $p(\omega_q)$ is the frequency of occurrence for class ω_q.

H.3 EXAMPLE

To illustrate the histogram analysis and segmentation technique, the image in Fig. H.2-5 was segmented. The zero-crossing analysis in Table H.3-1 reveals the presence of 7 modes at a scale of $\sigma = 2.5$. The mixture parameters estimated by this technique are listed in Table H.3-2.

Figure H.3-1 shows distributions due to water, crops, and other surface materials identified on the basis of their spectral reflectance in band 4. In Fig. H.3-2 relatively dark regions which contain pixels belonging to modes one and two are shown to correspond to bodies of water in Fig. H.2-5. In Fig. H.3-3 relatively bright regions which contain pixels belonging to modes five through seven are shown to correspond to crop fields and areas of denser vegetation.

TABLE H.3-1
ANALYSIS OF IMAGE HISTOGRAM

Lower Turning Point at 10 Peak at 13	Upper Turning Point at 87 Valley at 118
Upper Turning Point at 15 Valley at 20	Lower Turning Point at 120 Peak at 123
Lower Turning Point at 21 Peak at 22	Upper Turning Point at 130 Valley at 135
Upper Turning Point at 25 Valley at 30	Lower Turning Point at 140 Peak at 143
Lower Turning Point at 40	Upper Turning Point at 150 Valley at 177
Upper Turning Point at 46	Lower Turning Point at 188 Peak at 190
Lower Turning Point at 69 Peak at 77	Upper Turning Point at 193

TABLE H.3-2
ESTIMATED PARAMETERS OF MIXTURE

MODE	RELATIVE FREQUENCY	MEAN	STANDARD DEVIATION
1	0.055376	12.351486	0.638134
2	0.021731	21.138792	4.494987
3	0.046624	43.260418	7.069894
4	0.767219	77.959785	11.826522
5	0.076797	117.040428	12.510215
6	0.032249	142.022644	6.025322
7	0.000004	190.000000	1.000000

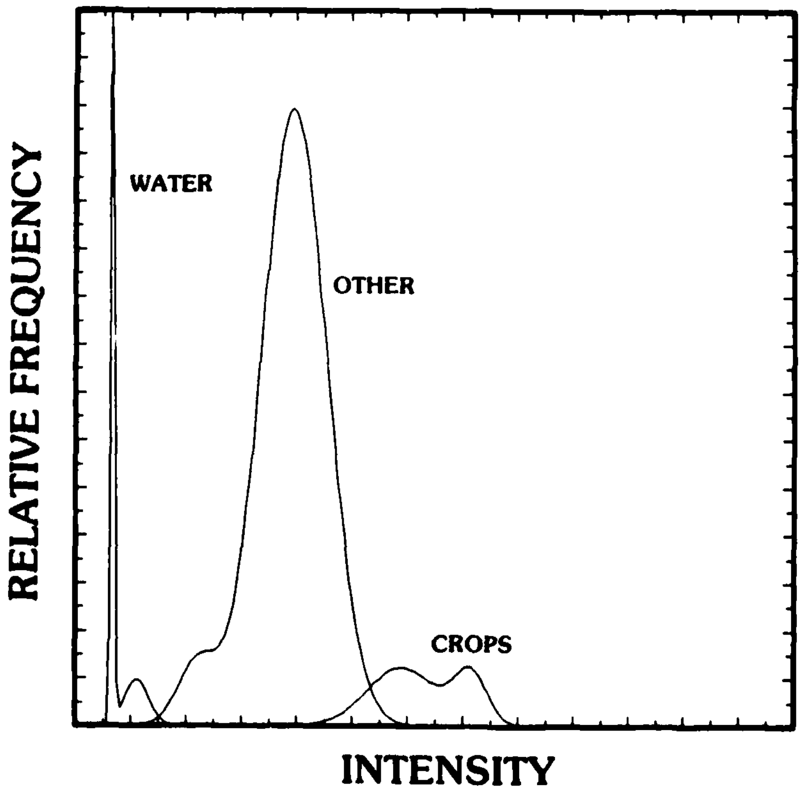


Figure H.3-1 Band 4 Histogram with Distributions Due to Water, Crops, and Other Materials Identified

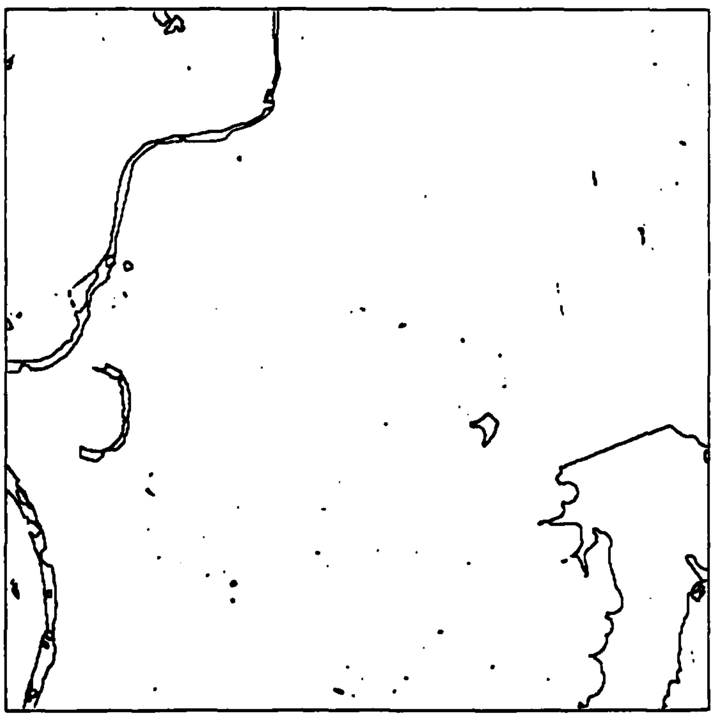


Figure H.3-2 Bodies of Water

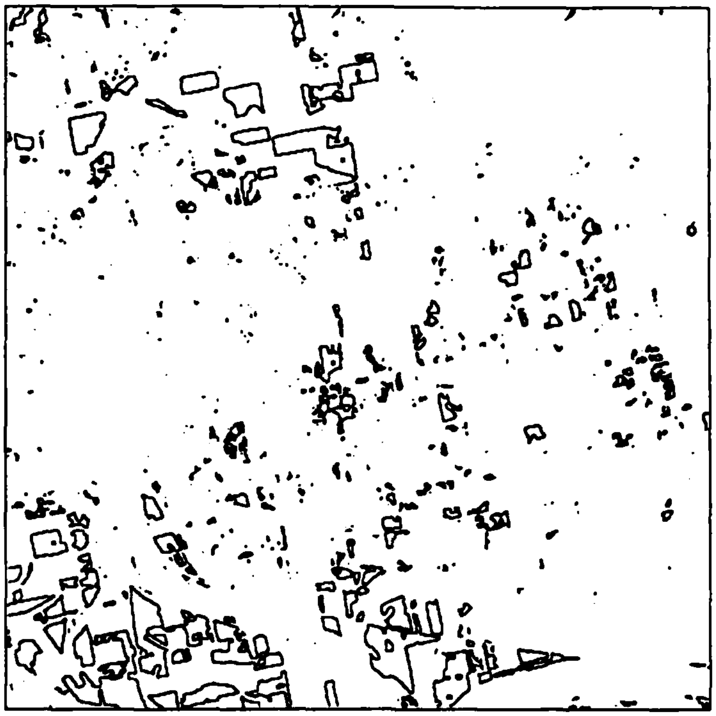


Figure H.3-3 Crops and Other Areas of Dense Vegetation

H.4 SUMMARY

A novel image segmentation technique was described in this appendix which involves analyzing the zero-crossings of derivatives of the image histogram convolved with a Gaussian, and estimating the parameters of an underlying mixture model. The salient features of the technique are that it does not require the presence or formation of well-defined peaks in the histogram to separate the individual modes in the data, it determines the number of modes and provides a maximum likelihood estimate of the mixture parameters for use in segmenting the image, and is computationally feasible. Preliminary results indicate that the segmentations obtained do correspond to physically-distinct regions in an image. An example of its use in extracting regions composed of water and crops based on their spectral reflectance in Landsat TM band 4 was presented. The example illustrates the manner in which it is used within the knowledge-based surface material classifier discussed in Appendix G.

REFERENCES

1. Laws, K.I., "On the Evaluation of Scene Analysis Algorithms," Image Understanding Workshop, 23 June 1983.

2. Defense Mapping Agency Pilot Digital Operations, "Automated Pattern Recognition for MC&G Features," Experiment 8, January 1983.

3. Kanade, T., "Region Segmentation: Signal vs Semantics," Computer Graphics and Image Processing, Vol. 13, 1980.

4. DLMS Level V Specification, Prototype Product Specification to Support High Resolution/Terrain Analysis Applications, Defense Mapping Agency, May 1982.

5. Nevatia, R., Machine Perception, Prentice-Hall, Inc., New Jersey, 1982.

6. Barrow, H.G. and Tenenbaum, J.M., "Computational Vision," Proceedings of the IEEE, Vol. 69, No. 5, May 1981.

7. Cohen, P.R. and Feigenbaum, E.A., The Handbook of Artificial Intelligence, William Kaufmann, Inc., CA, 1982.

8. Ballard, D.H. and Brown, C.M., Computer Vision, Prentice Hall, Inc., New Jersey, 1982.

9. Pratt, W.K., Digital Image Processing, John Wiley & Sons, New York, 1978.

10. Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, A Wiley-Interscience Publication, New York, 1973.

11. Tou, J.T. and Gonzalez, R.C., Pattern Recognition Principles, Addison-Wesley Publishing Co., Reading, MA, 1981.

12. Roberts, L.G., "Machine Perception of Three-dimensional Solids," in Optical and Electro-Optical Information Processing, J.P. Tippett et al. (eds.) MIT Press, Cambridge, MA, 1965.

13. Sobel, I., "Camera Models and Machine Perception," AIM-21, Stanford AI Lab, May 1970.

REFERENCES (Continued)

14. Prewitt, J.M.S., "Object Enhancement and Extraction," in Picture Processing and Psychopictorics, B.S. Lipkin and A. Rosenfeld (eds), Academic Press, New York, 1970.

15. Marr, D. and Hildreth, E., "Theory of Edge Detection," Proc. R. Soc. Lond. B., 207, 1980.

16. Herskovitz, A. and Binford, T.O., "On Boundary Detection," AI Memo 183, Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, 1970.

17. Rosenfeld, A., Thurston, M. and Lee, Y.H., "Edge and Curve Detection: Further Experiments," IEEE Trans. on Computers, Vol. C-21, No. 7, July 1982.

18. Kelly, M.D., "Edge Detection in Pictures by Computer Using Planning," in Machine Intelligence VI, Edinburgh University Press, Edinburgh, 1971.

19. Nevatia, R. and Babu, K.R., "Linear Feature Extraction and Description," Computer Graphics and Image Processing, Vol. 13, 1980.

20. Duda, R.O. and Hart, P.E., "Use of the Hough Transformation to Detect Lines and Curves in Pictures," Computer Graphics and Image Processing, Vol. 15, No. 1, January 1972.

21. Abdou, I.E. and Pratt, W.K., "Quantitative Design and Evaluation of Enhancement/Thresholding Edge Detectors," Proceeding of the IEEE, Vol. 67, No. 5, May 1978.

22. Modestino, J.W. and Fries, R.W., "Edge Detection in Noisy Images Using Recursive Digital Filtering," Computer Graphics and Image Processing, Vol. 6, 1977.

23. Montanari, U., "On the Optimal Detection of Curves in Noisy Pictures," Scientific Applications, Vol. 14, No. 5, May 1971.

24. Zucker, S.W., "Region Growing: Childhood and Adolescence," Computer Graphics and Image Processing, Vol. 5, 1976.

25. Risemand, E.M. and Arbib, M.A., "Computational Techniques in the Visual Segmentation of Static Scenes," Computer Graphics and Image Processing, Vol. 6, 1977.

REFERENCES (Continued)

26. Hanson, A.R. and Riseman, E.M., "VISIONS: A Computer System for Interpreting Scenes," in Computer Vision Systems, A.R. Hanson and E.M. Riseman (eds), Academic Press, NY, 1977.

27. Brice, C. and Fennema, C., "Scene Analysis Using Regions," Artificial Intelligence, 1, 1970.

28. Yakimovsky, Y. and Feldman, J., "A Semantics-Based Decision Theory Region Analyzer," Proc. Third Int. Joint Conf. on Artificial Intelligence, 1973.

29. Tanimato, S. and Pavlidis, T., "A Hierarchial Data Structure for Picture Processing," Computer Graphics and Image Processing, Vol. 4, 1975.

30. Ohlander, R., Price, K. and Reddy, D.R., "Picture Segmentation Using a Recursive Region Splitting Method," Computer Graphics and Image Processing, Vol. 8, 1978.

31. Coleman, G.B., "Image Segmentation by Clustering," Proceedings of the IEEE, Vol. 67, No. 5, May 1979.

32. Robertson, T.V., Swain, P.H. and Fu, K.S., "Multispectral Image Partitioning," TR-EE 73-26 (LARS Information Note 071373) School of Electrical Engineering, Purdue Univ., August 1973.

33. Laws, K.I., "Textured Image Segmentation," University of Southern California Report USCIPI 940 (Ph.D. thesis), January 1980.

34. Haralick, R.M., Statistical and Structural Approaches to Texture," Proceedings of the IEEE, Vol. 67, 1979.

35. Witkin, A.P., "Intensity-Based Edge Classification," Fairchild Laboratory for Artificial Intelligence Research, The National Conference on Artificial Intelligence, Carnegie-Mellon University, August 1982.

36. Pentland, A., "Fractal-based Description of Natural Scenes," Proceedings: Image Understanding Workshop, June 1983.

37. Lowe, D.G. and Binford, T.O., "The Perceptual Organization of Visual Images: Segmentation as a Basis for Recognition," Proceedings: Image Understanding Workshop, June 1983.

REFERENCES (Continued)

38. Fischler, M.A. and Bolles, R.C., "Perceptual Organization and Curve Partitioning," Proceedings: Image Understanding Workshop, June 1983.

39. Horn, B.K.P., "Obtaining Shape from Shading Information," in The Psychology of Computer Vision, P.H. Winston (ed), McGraw-Hill, 1975.

40. Horn, B.K.P., "Understanding Image Intensities," Artificial Intelligence, Vol. 8, 1977.

41. Gonzalez and Wintz, Digital Image Processing, Addison Wesley, Reading, MA, 1981.

42. Rosenfeld, A. and Kak, A.D., Digital Picture Processing, Academic Press, NY, 1976.

43. Andrews, H.C., Tescher, A.G. and Kruger, R.P., "Image Processing by Digital Computers," IEEE Spectrum, Vol. 9, No. 7, July 1972.

44. Hall, E.L., et al., "A Survey of Preprocessing and Feature Extraction Techniques for Radiographic Images," IEEE Trans. Computers, Vol. C-20, No. 9.

45. Hall, E.L., "Almost Uniform Distributions for Computer Image Enhancement," IEEE Trans. Computers, Vol. C-23, No. 2.

46. Ketcham, D.J., "Real Time Image Enhancement Techniques," Proceedings SPIE/OSA Conference on Image Processing (Pacific Grove, CA), Vol. 74, February 1976.

47. Frei, W., "Image Enhancement by Histogram Hyperbolization," Computer Graphics and Image Processing, Vol. 6, June 1977.

48. Hummel, R.A., "Histogram Modification Techniques," Computer Graphics and Image Processing, Vol. 4, September 1975.

49. Hummel, R.A., "Image Enhancement by Histogram Transformation," Computer Graphics and Image Processing, Vol. 6, April 1977.

50. Schreiber, W.F., "Image Processing for Quality Improvement," Proc. IEEE, Vol. 66, No. 12, December 1978.

REFERENCES (Continued)

51. Green, W.B., Digital Image Processing: A Systems Approach, Van Nostrand Reinhold, NY, 1983.

52. Peli, T. and Lim, J.S., "Adaptive Filtering for Image Enhancement," Proc. ICASSP-81 (Atlanta), Vol. 3, 1981.

53. Tom, V.T. and Wolfe, G.J., "Adaptive Histogram Equalization and its Applications," Proc. SPIE (San Diego), Vol. 359, August 1982.

54. O'Handley, D.A. and Green, W.B., "Recent Developments in Digital Image Processing at the Image Processing Laboratory at the Jet Propulsion Laboratory," Proc. IEEE 60.

55. Johnston, E.G. and Rosenfeld, A., "Geometrical Operations on Digitized Pictures," in Picture Processing and Psychopictorics, B.S. Lipkin and A. Rosenfeld, (eds.), Academic Press, NY, 1970.

56. Goldmark, P.C. and Hollywood, J.M., "A New Technique for Improving the Sharpness of Television Pictures," Proc. IRE 39, 1951.

57. Kovasznay, L.S.G. and Joseph, H.M., "Image Processing," Proc. IRE 43, 1955.

58. Arguello, R.J., Sellner, H.R. and Stuller, J.Z., "Transfer Function Compensation of Sampled Imagery," IEEE Trans. Comput., Vol. C-21, 1972.

59. Schreiber, W.F., "Wirephoto Quality Improvement by Unsharp Masking," Pattern Recognition, Vol. 2, No. 4, May 1970.

60. Stockham, T.G., Jr., "Image Processing in the Context of a Visual Model," Proc. IEEE 60, 1972.

61. Davis, L.S., "A Survey of Edge Detection Techniques," Computer Graphics and Image Processing, 1975.

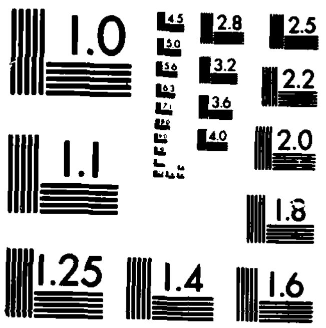
62. Peli, T. and Malah, D., "A Study of Edge Detection Algorithms," Computer Graphics and Image Processing, 1981.

63. Fries, R.W. and Modestino, J.W., "An Empirical Study of Selected Approaches to the Detection of Edges in Noisy Digital Images: TR77-1 Electrical System Engineering Dept., RPI, 1977.

AD-A150 189 FEATURE EXTRACTION ASSESSMENT STUDY(U) ANALYTIC SCIENCES CORP READING MA M J CARLOTTO ET AL. NOV 84 ETL-0377 DACA76-82-C-0004 4/4

UNCLASSIFIED F/G 8/2 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

REFERENCES (Continued)

64. Shaw, G.B., "Local and Regional Edge Detectors: Some Comparisons," Computer Graphics and Image Processing, 1979.

65. Fram, F.R. and Deutsch, E.S., "On the Quantitative Evaluation of Edge Detection Schemes and Their Comparison with Human Performance," IEEE Trans. Comp., Vol. C-24, No. 6, pp 616-628, June 1975.

66. Robinson, G.S., "Edge Detection by Compass Gradient Masks," Computer Graphics and Image Processing, 1977.

67. Rosenfeld, A. and Thurston, M., "Edge and Curve Detection for Visual Scene Analysis," IEEE Trans. Comp., Vol. C-20, No. 5, May 1971.

68. Hueckel, M.H., "An Operator which Locates Edges in Digital Pictures," J. Assoc. Comput. Mach., Vol. 18, No. 1, January 1971.

69. Hueckel, M.H., "A Local Visual Operator which Recognizes Edges and Lines," J. Assoc. Comput. Mach., Vol. 20, No. 4, October 1973.

70. Frei, W. and Chen, C.C., "Fast Boundary Detection: A Generalization and a New Algorithm," IEEE Trans. Comput., Vol. C-26, October 1977.

71. Jacobus, C.J. and Chien, R.T., "Two New Edge Detectors," T-PAMI 5, 1981.

72. Argyle, E., "Techniques for Edge Detection," Proc. IEEE 59, 1971.

73. MacLeod, I.D., "Comments on 'Techniques for Edge Detection'," Proc. IEEE 60, 1972.

74. Shanmugam, K.S., Dickey, F.M. and Green, J.A., "An Optimal Frequency Domain Filter for Edge Detection in Digital Pictures," T-PAMI 1, 1979.

75. Crochiere, R.E. and Rabiner, L.R., "Optimum FIR Digital Filter Implementation for Decimation, Interpolation and Narrow Band Filtering," IEEE Trans. Acoust. Speech Signal Proc., Vol. ASSP-23, No. 5.

REFERENCES (Continued)

76. Lee, J.S., "Digital Image Enhancement and Noise Filtering by Use of Local Statistics," T-PAMI 2, 1980.

77. Chin, R.T. and Yeh, C.L., "Quantitative Evaluation of Some Edge-preserving Noise-smoothing Techniques," Computer Graphics and Image Processing 23, 1983.

78. Tyan, S.G., "Median Filtering: Deterministic Properties," in Two-Dimensional Signal Processing: Transforms and Median Filters, (T.S. Huang, Ed.),Springer-Verlag, NY, 1981.

79. Nagao, M. and Matsuyama, T., "Edge Preserving Smoothing," Computer Graphics and Image Processing 9, 1979.

80. Davis, L.S. and Rosenfeld, A., "Noise Cleaning by Iterated Local Averaging," IEEE Trans. Syst. Man Cybern., Vol. SMC-8, 1978.

81. Dyer, C.R. and Rosenfeld, A., "Thinning Algorithms for Greyscale Pictures," T-PAMI 1, 1979.

82. Blum, H., "A Transformation for Extracting New Descriptions of Shape," Symposium on Models for the Perception of Speech and Visual Form, MIT Press, Cambridge, MA, 1964.

83. Montinari, U., "Continuous Skeletons from Digitized Images," J. ACM 16, 1969.

84. Yokoi, S., Toriwaki, J.I. and Fukumura, T., "Generalized Distance Transformation of Digital Binary Images," Proc. Fifth Intern. Conf. Pattern Recognition, December 1-4, 1980.

85. Pavlidis, T., "An Asynchronous Thinning Algorithm," Computer Graphics and Image Processing 20, 1982.

86. Pfaltz, J.L. and Rosenfeld, A., "Computer Representations of Planar Regions by their Skeletons," Commun. ACM 10, 1967.

87. Tamura, H., "A Comparison of Line Thinning Algorithms from a Digital Geometry Viewpoint," Proc. Fourth Intern. Joint Conf. Pattern Recognition, November 7-10, 1978.

REFERENCES (Continued)

88. Arcelli, C., Cordella, L.P. and Levialdi, "From Local Maxima to Connected Skeletons," IEEE Trans. Pattern Analysis Machine Intelligence, PAMI-3, 1981.

89. Hough, P.V., "Method and Means for Recognizing Complex Patterns," U.S. Patent 3,069,654, December 18, 1962.

90. Tannino, A. and Shapiro, S., "A Survey of the Hough Transform and Its Extensions for Curve Detection," Proc. IEEE Conf. on Pattern Recog. and Image Proc. (Chicago, IL), May 1978.

91. Martelli, A., "An Application of Heuristic Search Methods to Edge and Contour Detection," Comm. of ACM, Vol. 19, No. 2, 1976.

92. Ashkar, G. and Modestino, J., "The Contour Extraction Problem with Biomedical Applications," Computer Graphics and Image Processing, Vol. 7, No. 3, 1978.

93. Zucker, S., Hummel, R. and Rosenfeld, A., "An Application of Relaxation Labelling to Line and Curve Enhancement," IEEE Trans. Computers, Vol. 26, 1977.

94. Nakagawa, Y. and Rosenfeld, A., "Edge/border Coincidence as an Aid in Edge Extraction," Univ. of Md Comp. Sci. Ctr., TR-647, March 1978.

95. Brooks, R.A., "Goal-directed Edge Linking and Ribbon Finding," Proceedings ARPA Image Understanding Workshop, Menlo Park, April 1979, pp 72-76.

96. Muerle, J.L. and Allen, D.C., "Experimental Evaluation of Techniques for Automatic Segmentation of Objects in a Complex Scene," in Pictorial Pattern Recognition, G.C. Cheng et al (eds.),Thompson, Washington, 1968.

97. Feldman, J.A. and Yakimousky, Y., "Decision Theory and Artificial Intelligence: I.A Semantics-based Region Analyzer," Artificial Intelligence, Vol. 5, No. 4, 1974.

98. Horowitz, S.L. and Pavlidis, T., "Picture Segmentation by a Directed Split-and-Merge Procedure," Proceedings of the Second Int. Joint Conf. on Pattern Recognition, August 1974.

REFERENCES (Continued)

99. Chow, C.K. and Kaneko, T., "Automatic Boundary Detection of the Left Ventricle from Cineangeograms," Computers and Biomedical Research, Vol. 5, No. 4, August 1972.

100. Therrien, C.W., "Linear Filtering Models for Texture Classification and Segmentation," Proceedings of the Fifth Int. Conf. on Pattern Recognition, December 1980.

101. Shen, H.C. and Wong, K.C., "Generalized Texture Representation and Metric," Proceedings of the Int. Conf. on Cybernetics and Society, October 1980.

102. Bajcsy, R. and Lieberman, L., "Texture Gradient as a Depth Cue," Computer Graphics and Image Processing, Vol. 5, No. 1, March 1976.

103. Lim, J.S. and Malik, N.A., "A New Algorithm for Two-dimensional Maximum Entropy Power Spectral Estimation," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-29, No. 3, June 1981.

104. Chen, C.H. and Young, G.K., "On a Two-dimensional Maximum Entropy Spectral Estimation Method for Texture Image Analysis," SMU-EE-TR-81-16, SMU, N. Dartmouth, Mass., October 1981.

105. Tomita, F., Yachida, M. and Tsuji, S., "Detection of Homogenous Regions by Structural Analysis," Proceedings of the Third Int. Joint Conf. on Artificial Intelligence, Vol. 3, 1973.

106. Fu, K.S., Syntactic Methods in Pattern Recognition, Academic Press, New York, 1974.

107. Fisher, G.L., Pollack, D.K., Radack, B. and Stevens, M.E. (eds.), Optical Character Recognition, Spartan, Washington, D.C.

108. Program Management Information System, DMIS/P Functional Description Document, Defense Mapping Agency, June 1982.

109. Grimson, W.E.L., "From Images to Surfaces," MIT Press, 1981.

110. Nagao, M. and Matsuyama, T., "A Structural Analysis of Complex Aerial Photographs, Plenum Press, 1980.

REFERENCES (Continued)

111. Childers, D.G., Editor, "Modern Spectrum Analysis," IEEE Press, IEEE, Inc., New York, 1978.

112. Moik, J.G., "Digital Processing of Remotely Sensed Images," NASA SP-431, NASA Scientific and Technical Information Branch, Washington, DC, 1980.

113. Bernstein, R., Lotspiech, J.B., Myers, H.J., Kolsky, H.G., and Lees, R.D., "Analysis and Processing of LANDSAT-4 Sensor Data Using Advanced Image Processing Techniques and Technologies," IEEE Transactions on Geosciences and Remote Sensors, Vol. GE-22, No. 3, 1984, pp. 192-221.

114. Rindfleisch, T.C., et al., "Digital Processing of the Mariner 6 and 7 Pictures," Journal of Geophysics Research, Vol. 76, 1971, pp. 394-417.

115. Horn, B.K. and Woodham, R.J., "Destriping Satellite Images," Artificial Intelligence Laboratory Report AI 467, MIT, Cambridge, MA, 1978.

116. Goetz, A.F., Billingsley, F.C., Gillispie, A.R., Abrams, M.J., and Squires, R.L., "Application of ERTS Images and Image Processing to Regional Geologic Problems and Geologic Mapping in Northern Arizona,' NASA/JPL TR 32-1597, May 1975.

117. Quatieri, T.F., "Object Detection by Two-Dimensional Linear Prediction," MIT Lincoln Laboratory Technical Report 632, January 1983.

118. Pravdo, S.H., Huneycutt, B., Holt, B.M., and Held, D.N., "SEASAT Synthetic-Aperture Radar Data User's Manual," NASA/JPL Technical Report 82-90, March 1983.

119. Wie, P.V. and Stein, M., "A Landsat Digital Image Rectification System," IEEE Transactions on Geoscience Electronics, Vol. GE-15, No. 3, 1977, pp. 130-137.

120. Maurer, H.E., Oberholtzer, J.D., and Anuta, P.E., Editors, "Synthetic Aperture Radar/LANDSAT MSS Image Registration," NASA Reference Pub. 1039, June 1979.

121. Welch, R. and Usery, E.L., "Cartographic Accuracy of LANDSAT-4 MSS and TM Image Data," IEEE Transactions on Geoscience and Remote Sens., Vol. GE-22, No. 3, 1984, pp. 281-288.

REFERENCES (Continued)

122. Wecksung, G.W. and Breedlove, J.R., "Some Techniques for Digital Processing, Display and Interpretation of Radio Images in Multi-Spectral Remote Sensing," Proceedings, SPIE (San Diego), Vol. 119, August 1977, pp. 47-54.

123. Kauth, R.J. and Thomas, G.S., "The Tasseled Cap - A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT," Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University, W. Lafayette, IN, pp. 4841-4851.

124. Vogel, M.A., Wong, A.K.C., "PFS Clustering Method," IEEE Trans. Pattern Analysis and Machine Intelligency, PAMI-1, No. 3, July 1979.

125. Prewitt, J.S.M. and Mendelsohn, M.L., "The Analysis of Cell Images," Ann. NY Academic Science 128, 1966, pp. 1035-1053.

126. Lennington, R.K. and Sorensen, C.T., "A Mixture Model Approach for Estimating Crop Areas From LANDSAT Data," Remote Sensing of Environment, Vol. 14, 1984, pp. 197-206.

127. Swain, P.H., "Fundamentals of Pattern Recognition in Remote Sensing," Remote Sensing: The Quantative Approach, McGraw Hill, 1978.

128. Foley, D.H., "Considerations of Sample and Feature Size," IEEE Trans. Information Theory, Vo. IT-18, pp 618-626, September 1972.

129. Landgrebe, D.A., "Analysis Technology for Land Remote Sensing," Proceedings IEEE, Vol. 69, No. 5, May 1981.

130. Kettig, R.L. and Landgrebe, D.A., "Classification of Multi-Spectral Image Data by Extraction and Classification of Homogeneous Objects," IEEE Transactions on Geoscience Electronics, Vol. GE-14, No. 1, 1976.

131. Chittineni, C.B., "Utilization of Spectral-Spatial Information in the Classification of Imagery Data," Computer Graphics and Image Processing, Vol. 16, 1981, pp. 305-340.

132. Henderson, R.G., "Signature Extension via the MASC Algorithm," IEEE Transactions on Geoscience Electronics, Vol. GE-14, No. 1, nuary 1976.

REFERENCES (Continued)

133. Pavlidis, T., "Algorithms for Graphics and Image Processing," Computer Science Press, 1982.

134. Winston, P.H. and Horn, B.K.P, "LISP," Addison-Wesley, 1981.

135. Pavlidis, T. and Feng, H., "Shape Discrimination," Syntactic Pattern Recognition Applications, Springer-Verlag, 1977.

136. Assada, H., Brady, M., "The Curvature Primal Sketch," AI Memo 758, MIT Artificial Intelligence Laboratory.

137. Zahn, C.T., Roskies, R.Z., "Fourier Descriptors for Plane Closed Curve," IEEE Trans. Computers, Vol. C-21, No. 3, March 1972.

138. Van Oeffelen, M.P., Vos, P.G., "An Algorithm for Pattern Description on the Level of Relative Proximity," Pattern Recognition, Vol. 16, No. 3, pp. 341-348, 1983.

139. Marr, D., "Analyzing Natural Images: A Computational Theory of Texture Vision," AI Memo 334, MIT Artificial Intelligence Laboratory.

140. Brooks, R.A., "Symbolic Reasoning Among 3-D Models and 2-D Images," Ph.D. Thesis, University Microfilms Internation 8124043, 1981.

141. Feigenbaum, E.A. and Cohen, P.R., The Handbook of Artificial Intelligence, Vol. 3, William Kaufmann, Inc., 1982.

142. IU Algorithm Report, Hughes Aircraft Company, May 1982.

143. Erman, L.D., Hayes-Roth, F., Lesser, V.R., Reddy, D.R., "The Hearsay-II Speech Understanding System Integrating Knowledge to Resolve Uncertainty," Computing Surveys, Vol. 12, No. 2, June 1980.

144. Colwell, R.N. (Editor and Chief), Manual of Remote Sensing Section Edition; American Society of Photogrammetry, Falls Church, VA, 1983.

145. Slater, P.N., Remote Sensing-Optics and Optical Systems, Addison-Wesley, Reading, MA, 1980.

REFERENCES (Continued)

146. Ulaby, F.T., Moore, R.K. and Fung, A.K., "Microwave Remote Sensing-Active and Passive, Volume 1 - Microwave Remote Sensing Fundamentals and Radiometry," Addison-Wesley, Reading, MA, 1981.

147. Imaging Radar Science Working Group, "The SIR-B Science Plan," JPL Publication 82-78, NASA Jet Propulsion Laboratory, Pasadena, CA, December 1982.

148. "Shuttle Imaging Radar-C (SIR-C), Executive Summary," JPL Publication 83-47, NASA Jet Propulsion Laboratory, Pasadena, CA, 1 July 1983.

149. Chaues, P.S., Jr., "Atmospheric, Solar, and M.T.F. Corrections for ERTS Digital Imagery," Proceedings of the American Society of Photogrammetry, 1975, pp. 69-69a.

150. Podwysock, M.H., Bender, L.U., Falcone, N., and Jones, O.D., "A Preliminary Evaluation of LANDSAT-4 Thematic Mapper Data for Their Geometric and Radiometric Accuracies," U.S. Geological Survey, Reston, VA, NTIS Document No. N83-32136, 1983.

151. Wrigley, R.C., et al., "Thematic Mapper Image Quality: Preliminary Results," Presented at IGARSS 1983, NASA Ames Research Center, Moffett Field, CA, NTIS Document No. N83-33284, 1983.

152. Card, D.H. and Wrigely, R.C., "Assessment of Thematic Mapper Band-to-Band Registration by the Block Correlation Method," NASA Ames Research Center, Moffett Field, CA, NTIS Document No. N83-33286, 1983.

153. Ulaby, F.T., Moore, R.K., and Fung, A.K., "Microwave Remote Sensing - Active and Passive, Volume 2 - Radar Remote Sensing and Surface Scattering and Emission Theory," Addison-Wesley, Reading, MA, 1982.

154. Anuta, P.E., "Spatial Registration of Multi-Spectral and Multi-Temporal Digital Imagery Using Fast Fourier Transform Techniques," IEEE Transactions on Geoscience Electronics, Vol. GE-8, No. 4, October 1970, pp. 353-368.

155. Witkin, A.P., "Scale-Space Filtering: A New Approach to Multi-Scale Description," Proceedings ICASSP, San Diego, CA, March 1984.

REFERENCES (Continued)

156. Duda, R.O., Hart, P.E., Pattern Classification and Scene Analysis, Wiley-Interscience, 1973.

157. Feldman, J.A., Yakimovky, Y., "Decision Theory and Artificial Intelligence: A Semantics-Based Region Analyzer," Artificial Intelligence Vol. 5, pp. 349-371, 1974.

158. Tou, J.T., Gonzalez, R.C., Pattern Recognition Principles, Addison-Wesley, 1974.

159. Tom, V.T., Wallace, G.K., and Wolfe, G.J., "Image Registration by a Statistical Method," Proc. SPIE, Vol. 432, San Diego CA, 1983.

160. Carlotto, M.J., "Texture Classification Based on Hypothesis Testing Approach," Seventh International Conference on Pattern Recognition, Montreal, Canada, 1982.

161. Wolfe, G.J., and Tom, V.T., "A VOD Production System Concept," TASC TR-4058-4, 5 June 1984.

162. Statistical Package for the Social Sciences, pp. 515-527, McGraw Hill, New York, 1975.

163. Tom, V.T., Carlotto, M.J., and Scholten, D.K., "Spatial Resolution Improvement of TM thermal Band Data," Proc. SPIE, Vol. 504, San Diego, CA, 1984.

END

FILMED

4-85

DTIC